

GENOMICS-ENABLED TRITICEAE IMPROVEMENT

EDITED BY: Xue-Feng Ma, Peter Stephen Baenziger, Shuyu Liu,
Xianchun Xia and Hakan Ozkan

PUBLISHED IN: Frontiers in Plant Science





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88974-890-7

DOI 10.3389/978-2-88974-890-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

GENOMICS-ENABLED TRITICEAE IMPROVEMENT

Topic Editors:

Xue-Feng Ma, Forage Genetics International West Salem, United States

Peter Stephen Baenziger, University of Nebraska System, United States

Shuyu Liu, Texas A&M University System, United States

Xianchun Xia, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, China

Hakan Ozkan, Çukurova University, Turkey

Citation: Ma, X.-F., Baenziger, P. S., Liu, S., Xia, X., Ozkan, H., eds. (2022). Genomics-enabled Triticeae Improvement. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-890-7

Table of Contents

- 05 Editorial: Genomics-Enabled Triticeae Improvement**
Xue-Feng Ma, Xianchun Xia, Shuyu Liu, Peter Stephen Baenziger and Hakan Özkan
- 08 Genetic Mapping by Integration of 55K SNP Array and KASP Markers Reveals Candidate Genes for Important Agronomic Traits in Hexaploid Wheat**
Hongchun Xiong, Yuting Li, Huijun Guo, Yongdun Xie, Linshu Zhao, Jiayu Gu, Shirong Zhao, Yuping Ding and Luxiang Liu
- 24 Genome-Wide Identification and Low Temperature Responsive Pattern of Actin Depolymerizing Factor (ADF) Gene Family in Wheat (*Triticum aestivum* L.)**
Ke Xu, Yong Zhao, Sihang Zhao, Haodong Liu, Weiwei Wang, Shuhua Zhang and Xueju Yang
- 38 The WtmsDW Locus on Wheat Chromosome 2B Controls Major Natural Variation for Floret Sterility Responses to Heat Stress at Booting Stage**
Million F. Erena, Iman Lohraseb, Isabel Munoz-Santa, Julian D. Taylor, Livinus C. Emebiri and Nicholas C. Collins
- 52 Molecular Cytogenetic and Agronomic Characterization of the Similarities and Differences Between Wheat–*Leymus mollis* Trin. and Wheat–*Psathyrostachys huashanica* Keng 3Ns (3D) Substitution Lines**
Jiachuang Li, Jiaojiao Li, Xueni Cheng, Li Zhao, Zujun Yang, Jun Wu, Qunhui Yang, Xinhong Chen and Jixin Zhao
- 68 Genomic Prediction of Grain Yield in a Barley MAGIC Population Modeling Genotype per Environment Interaction**
Damiano Puglisi, Stefano Delbono, Andrea Visioni, Hakan Ozkan, İbrahim Kara, Ana M. Casas, Ernesto Igartua, Giampiero Valè, Angela Roberta Lo Piero, Luigi Cattivelli, Alessandro Tondelli and Agostino Fricano
- 86 Identification, Characterization, and Expression Profile Analysis of the mTERF Gene Family and Its Role in the Response to Abiotic Stress in Barley (*Hordeum vulgare* L.)**
Tingting Li, Wenqiu Pan, Yiyuan Yuan, Ying Liu, Yihan Li, Xiaoyu Wu, Fei Wang and Licao Cui
- 106 The Haplotype-Based Analysis of *Aegilops tauschii* Introgression Into Hard Red Winter Wheat and Its Impact on Productivity Traits**
Moses Nyine, Elina Adhikari, Marshall Clinesmith, Robert Aiken, Bliss Betzen, Wei Wang, Dwight Davidson, Zitong Yu, Yuanwen Guo, Fei He, Alina Akhunova, Katherine W. Jordan, Allan K. Fritz and Eduard Akhunov
- 122 Multi-Trait Multi-Environment Genomic Prediction of Agronomic Traits in Advanced Breeding Lines of Winter Wheat**
Harsimardeep S. Gill, Jyotirmoy Halder, Jinfeng Zhang, Navreet K. Brar, Teerath S. Rai, Cody Hall, Amy Bernardo, Paul St Amand, Guihua Bai, Eric Olson, Shaukat Ali, Brent Turnipseed and Sunish K. Sehgal

- 136 Multiple Wheat Genomes Reveal Novel Gli-2 Sublocus Location and Variation of Celiac Disease Epitopes in Duplicated α -Gliadin Genes**
Gwyneth Halstead-Nussloch, Tsuyoshi Tanaka, Dario Copetti, Timothy Paape, Fuminori Kobayashi, Masaomi Hatakeyama, Hiroyuki Kanamori, Jianzhong Wu, Martin Mascher, Kanako Kawaura, Kentaro K. Shimizu and Hirokazu Handa
- 147 Genome-Wide Linkage Mapping for Preharvest Sprouting Resistance in Wheat Using 15K Single-Nucleotide Polymorphism Arrays**
Lingli Li, Yingjun Zhang, Yong Zhang, Ming Li, Dengan Xu, Xiuling Tian, Jie Song, Xumei Luo, Lina Xie, Desen Wang, Zhonghu He, Xianchun Xia, Yan Zhang and Shuanghe Cao
- 156 A Genome-Wide Association Study Pinpoints Quantitative Trait Genes for Plant Height, Heading Date, Grain Quality, and Yield in Rye (*Secale cereale* L.)**
Dörthe Siekmann, Gisela Jansen, Anne Zaar, Andrzej Kilian, Franz Joachim Fromme and Bernd Hackauf
- 179 Genome-Wide Identification of Wheat ZIP Gene Family and Functional Characterization of the TaZIP13-B in Plants**
Song Li, Zihui Liu, Linlin Guo, Hongjie Li, Xiaojun Nie, Shoucheng Chai and Weijun Zheng
- 197 Identification and Validation of Stable Quantitative Trait Loci for SDS-Sedimentation Volume in Common Wheat (*Triticum aestivum* L.)**
Shuai Tian, Minghu Zhang, Jinghui Li, Shaozhe Wen, Chan Bi, Huanhuan Zhao, Chaoxiong Wei, Zelin Chen, Jiazheng Yu, Xintian Shi, Rongqi Liang, Chaojie Xie, Baoyun Li, Qixin Sun, Yufeng Zhang and Mingshan You
- 208 Genome-Wide Association and Genomic Prediction for Stripe Rust Resistance in Synthetic-Derived Wheats**
Zahid Mahmood, Mohsin Ali, Javed Iqbal Mirza, Muhammad Fayyaz, Khawar Majeed, Muhammad Kashif Naeem, Abdul Aziz, Richard Trethowan, Francis Chuks Ogbonnaya, Jesse Poland, Umar Masood Quraishi, Lee Thomas Hickey, Awais Rasheed and Zhonghu He



Editorial: Genomics-Enabled Triticeae Improvement

Xue-Feng Ma^{1,2*}, Xianchun Xia³, Shuyu Liu⁴, Peter Stephen Baenziger⁵ and Hakan Özkan⁶

¹ Noble Research Institute, Ardmore, OK, United States, ² Forage Genetics International, West Salem, WI, United States,

³ Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, ⁴ Texas A&M AgriLife Research,

Amarillo, TX, United States, ⁵ Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE, United States,

⁶ Department of Field Crops, Faculty of Agriculture, University of Çukurova, Adana, Turkey

Keywords: abiotic stress, agronomic traits, barley, biotic stress, genomics, quality traits, rye, wheat

Editorial on the Research Topic

Editorial: Genomics-Enabled Triticeae Improvement

INTRODUCTION

Triticeae, an important tribe in the grass family Poaceae, includes several staple food crops, such as wheat, barley, rye and triticale. Although plant breeding has improved the performance of these crops steadily in agricultural traits, continued improvement is becoming more challenging due to increasing pressure from biotic and abiotic stresses as a result of climate change.

Further advance in structural and functional genomics of the Triticeae crops is essential for accelerating cycles of crop improvement and increasing genetic gain of breeding selection. Marker-assisted selection has proven to be useful for tracking alleles of major genes of many traits of agricultural importance. More recently, genomic selection has also shown potential for simplifying the selection of genome-wide minor alleles by modeling with or without pre-knowledge of the target traits. However, it has been noted that almost all Triticeae breeding programs still rely largely on conventional breeding selection made from replicated, time-consuming field trials.

With the recent advancements on reference sequences of the major Triticeae species and high-throughput genotyping platforms, we envision that more feasible genomic tools can be developed rapidly and more germplasm resources can be characterized precisely, thus, increasing the certainty of higher genetic gain through breeding. Therefore, this Research Topic aims to promote genomics-enabled Triticeae improvement by collecting original research articles, especially involving use of the most recent genomic resources in Triticeae crops.

We are honored to receive submissions of a large number of manuscripts addressing various subject areas across major Triticeae crops. After vigorous reviewing and revising, 14 of them were collected in this Research Topic, covering reports on yield components, grain quality traits, and tolerance to biotic/abiotic stresses in wheat, barley and rye.

Overall, quantitative trait loci (QTL) mapping remains a large area in this Research Topic for developing molecular breeding tools. We were also glad to receive contributions on genomic prediction modeling, especially involving multivariate prediction models covering multi-traits and multi-environments with various cross-validation schemes. These studies used different types of experimental populations such as biparental, multi-parent advanced generation intercrosses (MAGIC) or genome-wide association study (GWAS) populations. In addition, the current Research Topic also included novel genomic resources, such as alien chromosome introgression or

OPEN ACCESS

Edited and reviewed by:

Yong-Bi Fu,
Agriculture and Agri-Food
Canada, Canada

*Correspondence:

Xue-Feng Ma
xma@foragegenetics.com

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 08 February 2022

Accepted: 22 February 2022

Published: 18 March 2022

Citation:

Ma X-F, Xia X, Liu S, Baenziger PS
and Özkan H (2022) Editorial:
Genomics-Enabled Triticeae
Improvement.
Front. Plant Sci. 13:871816.
doi: 10.3389/fpls.2022.871816

substitution lines. The studies characterized novel favorable alleles that are important for wheat improvement. Furthermore, a few studies reported molecular characterization of transcription factors and transporters associated with plant development and response to diverse growing conditions. These articles are highlighted below according to major traits being studied.

AGRONOMIC TRAITS

The improvement of Triticeae crops has been continuously benefited from genetic introgression of wild relatives, which provide a great potential to broaden the availability of favorable genetic alleles that are otherwise not available in the primary gene pools of the crops. In this Research Topic, Nyine et al. assessed the impact of the introgression from 21 diverse accessions of *Aegilops tauschii*, the diploid ancestor of the wheat D genome. Using whole-genome sequencing of parental lines and the sequence-based genotyping of an BC₁F_{3:5} introgression population together with phenotyping data collected from field trials, they revealed some introgression lines that could increase grain yield. They also identified SNPs and haplotypes that were significantly associated with yield component traits and genes regulating plant development. The study provided valuable germplasm with characterized haplotypes of *Ae. tauschii* for wheat improvement.

Advancing molecular breeding tools has been a continuous task facilitating crop development. Xiong et al. located QTL associated with important agronomic traits in hexaploid wheat and developed diagnostic kompetitive allele-specific PCR (KASP) markers for the traits to facilitate wheat breeding.

In this Research Topic, agronomic traits were also mapped in rye, which is the only cross-pollinating Triticeae crop species. Siekmann et al. reported the first GWAS of agronomic traits evaluated from experimental hybrids of rye, and located cross-validated SNPs in protein-coding genes associated with plant height, heading date, grain quality and yield.

Besides QTL mapping, this Research Topic also covered genomic predication studies in wheat and barley. The winter wheat study by Gill et al. used multivariate genomic prediction models to predict several agronomic traits using advanced and elite breeding lines evaluated in multiple environments. They evaluated prediction accuracy of a multi-trait model with two cross-validation schemes and a multi-trait multi-environment model that integrates the analysis of multiple traits. Results showed that multivariate genomic selection models have great potential in implementing genomic selection in breeding programs.

In barley, genomic prediction for grain yield was modeled by Puglisi et al. using a MAGIC population derived from eight founders. Predictive abilities were evaluated for single-environment genomic prediction and multi-environment genomic prediction models with various cross-validation schemes. The study concluded, in general, multi-environment models that explicitly split marker effects in main and environmental-specific effects outperform simpler multi-environment models.

GRAIN QUALITY TRAITS

The Research Topic also includes studies on improving grain quality traits. Tian et al. mapped QTL for sodium dodecyl sulfate (SDS)-sedimentation volume (SSV), an important index for gluten strength of common wheat. Notably, environmentally stable QTL were detected and additive effect of two closely linked QTL on chromosome 1A was illustrated. They also characterized favorable loci for improving SSV, and proposed an ideal target for positional cloning.

In addition, Li L. et al. located QTL underneath wheat preharvest sprouting (PHS), which significantly reduces grain yield and quality. The research not only provided genetic resources for PHS resistance but also developed KASP markers tightly linked to germination index for marker-assisted breeding.

Furthermore, Halstead-Nussloch et al. revealed a novel *Gli-2* sublocus using 11 recently published chromosome-scale assemblies of hexaploid wheat. The research analyzed genomic variation in α -gliadins and unexpectedly found that the *Gli-B2* locus comprises two subloci. The research also confirmed variation of celiac disease epitopes in duplicated α -gliadin genes. The analysis yielded a new pass for improving grain quality through wheat breeding.

BIOTIC STRESS

Exploiting resistant resources from wild relatives has played critical roles in coping with various stresses in cereal crops. Enclosed in this Research Topic, Li J. et al. characterized wheat–*Leymus mollis* Trin. and wheat–*Psathyrostachys huashanica* Keng 3Ns (3D) substitution lines. The characterization generated new genetic resources for disease resistance and high-yield breeding with characterized substitution lines showing superior resistance to powdery mildew or *Fusarium* head blight.

Advancement in stripe rust resistance was also reported in this Research Topic. Using synthetic-derived wheats, Mahmood et al. located a large number of quantitative trait nucleotides, including some novel loci and haplotypes from *Ae. tauschii*. They also evaluated different models for genomic prediction of stripe rust resistance, and reported encouraging prediction accuracy for adult-plant resistance to stripe rust.

ABIOTIC STRESS

Abiotic stress is an increasing challenge in cereal production. Heat stress at booting stage causes significant losses to floret fertility (grain set) and hence yield in wheat. Erena et al. identified a major-effect heat tolerance locus on wheat chromosome 2B. The locus offsets between 44 and 65% of the losses in grain set due to heat, suggesting that it offers significant value for marker-assisted wheat breeding against heat stress.

In an effort to address wheat cold stress, Xu et al. identified genome-wide actin depolymerizing factor (ADF) genes, and characterized them using transgenic analysis. The effort generated fundamental information about the wheat ADF genes, their potential regulatory effects of the

encoded proteins on plant development and responses to low-temperature stress.

In addition, Li S. et al. investigated the wheat ZIP (Zn-regulated, iron-regulated transporter-like protein) transporter, which plays an important role in regulating the uptake, transport, and accumulation of microelements in plants. The investigation searched ZIP genes against the wheat reference genome and then systematically analyzed the gene structure, expression profiles, regulatory network, and biological function regulating stress responses to microelements.

Abiotic stress was also investigated in barley. Li T. et al. studied plant mitochondrial transcription termination factor (mTERF) family, which regulates organellar gene expression. Expression analysis suggested that some members of the mTERF family were significantly induced by various abiotic stresses or phytohormone treatment, suggesting their important roles in regulating stress responses.

Altogether, the range of research in this topic clearly illustrates diverse efforts on improving Triticeae crops although the articles in this Research Topic are still very limited. The topic also highlighted the importance of public genomic resources given the fact that most of the studies involved usage of public reference genomes.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

We are grateful to all authors, journal editors, and peer reviewers who contributed to this Research Topic.

Conflict of Interest: X-FM was employed by Forage Genetics International.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ma, Xia, Liu, Baenziger and Özkan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genetic Mapping by Integration of 55K SNP Array and KASP Markers Reveals Candidate Genes for Important Agronomic Traits in Hexaploid Wheat

Hongchun Xiong[†], Yuting Li[†], Huijun Guo[†], Yongdun Xie, Linshu Zhao, Jiayu Gu, Shirong Zhao, Yuping Ding and Luxiang Liu*

National Engineering Laboratory for Crop Molecular Breeding, National Center of Space Mutagenesis for Crop Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

OPEN ACCESS

Edited by:

Peter Stephen Baenziger,
University of Nebraska System,
United States

Reviewed by:

Waseem Hussain,
International Rice Research Institute
(IRRI), Philippines
Shantel Amealia Martinez,
Washington State University,
United States

*Correspondence:

Luxiang Liu
liuluxiang@caas.cn

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 12 November 2020

Accepted: 29 January 2021

Published: 23 February 2021

Citation:

Xiong H, Li Y, Guo H, Xie Y,
Zhao L, Gu J, Zhao S, Ding Y and
Liu L (2021) Genetic Mapping by
Integration of 55K SNP Array
and KASP Markers Reveals
Candidate Genes for Important
Agronomic Traits in Hexaploid Wheat.
Front. Plant Sci. 12:628478.
doi: 10.3389/fpls.2021.628478

Agronomic traits such as heading date (HD), plant height (PH), thousand grain weight (TGW), and spike length (SL) are important factors affecting wheat yield. In this study, we constructed a high-density genetic linkage map using the Wheat55K SNP Array to map quantitative trait loci (QTLs) for these traits in 207 recombinant inbred lines (RILs). A total of 37 QTLs were identified, including 9 QTLs for HD, 7 QTLs for PH, 12 QTLs for TGW, and 9 QTLs for SL, which explained 3.0–48.8% of the phenotypic variation. Kompetitive Allele Specific PCR (KASP) markers were developed based on sequencing data and used for validation of the stably detected QTLs on chromosomes 3A, 4B and 6A using 400 RILs. A QTL cluster on chromosome 4B for PH and TGW was delimited to a 0.8 Mb physical interval explaining 12.2–22.8% of the phenotypic variation. Gene annotations and analyses of SNP effects suggested that a gene encoding protein Photosynthesis Affected Mutant 68, which is essential for photosystem II assembly, is a candidate gene affecting PH and TGW. In addition, the QTL for HD on chromosome 3A was narrowed down to a 2.5 Mb interval, and a gene encoding an R3H domain-containing protein was speculated to be the causal gene influencing HD. The linked KASP markers developed in this study will be useful for marker-assisted selection in wheat breeding, and the candidate genes provide new insight into genetic study for those traits in wheat.

Keywords: QTL, heading date, plant height, thousand grain weight, spike length, wheat

INTRODUCTION

Wheat (*Triticum aestivum* L.) is one of the most important cereal crops worldwide, providing a food source for 30% of the human population (Mayer et al., 2014). Improving the yield potential of wheat is of great significance for meeting the food demand from an increasing population (Tshikunde et al., 2019). Agronomic traits such as heading date (HD), plant height (PH), thousand grain weight (TGW), and spike length (SL) are important factors affecting yield and always targeted by wheat breeders (Tshikunde et al., 2019). Recent advances in wheat genomics have accelerated the genetic

dissection of important agronomic traits, and a large number of quantitative trait loci (QTLs) for these traits have been identified (Rasheed and Xia, 2019).

Heading date is crucial for adaptation to different environments and yield stability in wheat (Snape et al., 2001). Over a hundred QTLs for HD located across all wheat chromosomes have been detected (Milec et al., 2014; Kiseleva and Salina, 2018). The cloned genes affecting HD or flowering in wheat are mainly classified into three groups: vernalization (*VRN*), photoperiod (*Ppd*), and earliness *per se* (*Eps*) genes (Snape et al., 2001). Four *VRN* genes (*VRN1*, *VRN2*, *VRN3*, and *VRN4*) located on chromosome 5 or 7 of the A/B/D genomes, have been identified by map-based cloning (Yan et al., 2003, 2004, 2006; Kippes et al., 2015; Xie et al., 2019). The *Ppd* genes for photoperiod responses in wheat are mainly located on chromosomes 2A, 2B, and 2D (Beales et al., 2007). The *Eps* genes were identified on chromosome 1A^m in *Triticum monococcum* (Alvarez et al., 2016) and on long arm of chromosome 1D in hexaploid wheat (Zikhali et al., 2014).

Plant height is another important factor affecting yield potential in wheat (Flintham et al., 1997). Twenty-five reduced height genes (*Rht*), *Rht1* to *Rht25*, have been identified in wheat (Mo et al., 2018). According to the distinct responses to exogenous gibberellic acid (GA), these *Rht* genes were classified into GA-sensitive or GA-insensitive categories (Lou et al., 2016). The “green revolution” genes *Rht-B1b* (*Rht1*) and *Rht-D1b* (*Rht2*) located on chromosome 4B and 4D, respectively, encode truncated DELLA proteins, which are involved in the gibberellin signaling pathway (Peng et al., 1999). *Rht4*, *Rht5*, *Rht7*, *Rht8*, *Rht9*, *Rht12*, *Rht13*, *Rht22*, and *Rht23* are located on 2B, 3B, 2A, 2D, 7B, 5A, 7B, 7A, and 5D, respectively (Peng et al., 1999; Ellis et al., 2005; Asplund et al., 2012; Chen et al., 2015; Vikhe et al., 2017). *Rht24* is located on 6AL (Tian et al., 2017; Wurschum et al., 2017) while *Rht14*, *Rht16*, *Rht18*, and *Rht25* are located on 6AS (Haq et al., 2011; Grant et al., 2018; Mo et al., 2018).

TGW is one of the three essential components of grain yield. Most of the cloned genes associated with TGW in wheat were identified using a homology-based strategy (Chen et al., 2020). The wheat *TaGL3-5A* gene has been cloned, and a SNP in the 11th exon of *TaGL3-5A* is associated with variation in grain length and TGW (Yang et al., 2019). In addition, the *TaGW2* gene in wheat is well studied for its function in regulating grain weight (Su et al., 2011; Bednarek et al., 2012; Jaiswal et al., 2015; Simmonds et al., 2016; Zhai et al., 2018; Zhang et al., 2018). Through genetic linkage analyses, stable QTLs explaining over 10% of the phenotypic variance for TGW were identified on chromosomes 1A (Varshney et al., 2000), 1B (Mir et al., 2012), 2D (Ma et al., 2019), 3A (Cui et al., 2014a), 3D (Cui et al., 2014a; Kumar et al., 2016), 4A (Araki et al., 1999), 4B (Kumar et al., 2016; Guan et al., 2018; Xu et al., 2019; Chen et al., 2020), 5A (Börner et al., 2002; Cuthbert et al., 2008; Mir et al., 2012; Kumar et al., 2016), 5B (Yang et al., 2020), 5D (Li et al., 2018), 6A (Mir et al., 2012), 6D (Cui et al., 2014a), 7A (Kumar et al., 2006, 2016; Mir et al., 2012), and 7D (Chen et al., 2020).

Spike architecture traits such as spike length (SL) are tightly related to grain production in wheat (Yao et al., 2019). A number of studies have identified stable QTLs for SL on chromosomes

1A, 1B, 2D, 3A, 3B, 4A, 4B, 5A, 5B, 6A, 6B, 6D, 7A, 7B, and 7D (Li et al., 2002, 2018; Marza et al., 2006; Deng et al., 2011; Cui et al., 2012a; Liu et al., 2018a; Cao et al., 2019; Chai et al., 2019; Wolde et al., 2019; Yao et al., 2019; Hu et al., 2020). It has been reported that the *Q* gene on chromosome 5A, which encodes an AP2 transcription factor, affects SL in wheat (Kawaura et al., 2009).

To obtain the genetic basis for HD, PH, TGW, and SL, we conducted QTL mapping based on a RIL population in the present study. In our previous study we used Bulk Segregant Analysis (BSA) and identified *VRN-B1* as the gene responsible for HD variation in the RIL population (Li et al., 2020). In this study, we used the Wheat55K SNP Array to map QTLs for HD, PH, TGW, and SL in this RIL population. Moreover, we validated the major QTLs on chromosomes 3A, 4B, and 6A by developing Kompetitive Allele Specific PCR (KASP) markers based on sequencing data and predicted candidate genes for PH, TGW, and HD according to gene annotation and SNP effects analysis.

MATERIALS AND METHODS

Plant Materials and Phenotype Evaluation

As previously described (Li et al., 2020), a RIL population (400 lines) derived from a cross between an early heading mutant (*eh1*) and Lunxuan987 (LX987) was used for genetic mapping; generations F₆ to F₈ of the RIL population were included in this study. The RIL and parent lines were planted at the Zhongpuchang field station of the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (Beijing, China) during the 2015–2016, 2016–2017, and 2017–2018 cropping seasons. For each year, the experiment was conducted once and we selected three representative plants for phenotypic collection. A total of 15 plants for each line were planted in a row of 1 m, and the field conditions were managed according to local standard practices.

For HD, when more than half of the spikes had emerged from two thirds of the plants in a line, the date for that line was recorded (Li et al., 2020). At agronomic harvest maturity, three representative plants from the middle of each row showing uniform growth status were used for PH, TGW, and SL evaluation and the mean values from these three plants were used for QTL mapping. PH from each representative plant was measured from the ground to the tip of the spike excluding awns. After drying, the grain weight from each representative plant was measured and the number of grains was counted. TGW was calculated as the plant grain weight divided by the number of grains per plant multiplied by 1,000. SL from main stem of each representative plant was measured from the base of the rachis to the tip of the terminal spikelet excluding awns. The HD, PH, and TGW data were collected in 2016, 2017, and 2018, and the SL data were collected in 2016 and 2018. Analyses of variance, correlation coefficients, and broad sense heritability were performed using the ANOVA analysis tools of the QTL IciMapping v4.1 program¹.

¹<http://www.isbreeding.net/>

Genotyping

Genomic DNA of each RIL and parent line was extracted as previously described (Li et al., 2020). After assessment of DNA integrity and quantity, the DNA from 207 lines that were also used for KASP assay, along with the parent DNA samples were hybridized to the Wheat55K SNP Array containing 53,063 markers. The genotyping was performed by China Golden Marker (Beijing) Biotech Co. Ltd².

Genetic Map Construction and QTL Analysis

High quality genotyping data were obtained by filtering with a Dish QC threshold of >0.82 and a Call-Rate threshold of >95%. The BIN function of IciMapping 4.1 was used to remove redundant markers from poly-high-resolution (PHR) SNPs, and the SNPs with >25% missing data were filtered out. The genetic map was constructed by randomly selecting only one marker from each bin using the MAP function of IciMapping 4.1. The threshold of the logarithm of odds (LOD) score was set to 2.5, and the Kosambi map function was used to calculate the map distance from recombination frequencies. Composite interval mapping (ICIM) in IciMapping 4.1 was selected to identify QTLs for HD, PH, TGW, and SL. The mean values of phenotypic traits for each line in each cropping season were used for QTL analysis. QTL region was determined by the positions of left and right markers identified by IciMapping 4.1, and physical positions of markers on the wheat reference genome v1.0 are shown in **Supplementary Table 1**. QTLs for the same traits identified in 2 or 3 years were considered to be stable. Multi-Environment Traits (MET) analysis of QTL IciMapping v4.1 was used for assessment of QTL × environment interactions (Li et al., 2015).

Development of KASP Markers and QTL Validation

According to the SNPs between *eh1* and LX987 identified by RNA sequencing (RNA-seq) (Li et al., 2020), KASP markers around or in the region of stable QTLs specific for different subgenomes were designed using the polyploid primer design pipeline PolyMarker³. After evaluation of the polymorphisms between two parent lines, the developed KASP markers were used for genotyping the entire mapping population. The successfully developed KASP markers are listed in **Supplementary Table 2**. A total of 400 RILs were genotyped with KASP markers on chromosomes 3A, 4B, and 6A. The reaction volume and PCR procedures for the KASP assay were as previously described (Li et al., 2020), and the CFX 96 Real-Time System (Bio Rad, Hercules, CA, United States) was used for PCR and data analysis. QTL analysis was conducted using IciMapping 4.1.

Analysis of SNP Effects and Prediction of Candidate Genes

Based on RNA-seq data, which was collected from young spikes of *eh1* and LX987 when *eh1* was beginning to head (Li et al., 2020),

the SNPs between *eh1* and LX987 covering the intervals of flanking markers from QTL validation were obtained for SNP effects analysis. SNP effects were analyzed by Python⁴ according to the example and scripts from the website⁵. A score for missense variation is generated that reflects the predicted effect of the SNP on gene function. The more negative a score, the larger the effects on gene function. The SNPs with larger effects on gene function were speculated to be located in the candidate genes. Gene functions were predicted by searching for homologous genes in rice (*Oryza sativa*) and *Arabidopsis thaliana* using the Triticeae Multi-omics Center website⁶.

RESULTS

Phenotypic Variation in the RIL Population

Our previous study showed that there is variation in HD in a RIL population of 400 lines derived from a cross between the early heading mutant *eh1* and LX987 (Li et al., 2020). In addition to HD, we also found that PH, TGW, and SL differed between *eh1* and LX987; the values of these traits were significantly lower in *eh1* than in LX987 from 2016–2018 (**Table 1**). Therefore, phenotypic investigation of PH, TGW, and SL in the RIL population was also conducted from 2016–2018. In the RIL population, the percent variation in PH, TGW, and SL ranged from 9.1% to 14.4% from 2016–2018, and all three traits showed moderate h^2 values ranging from 0.77 to 0.82 (**Table 1**).

⁴<https://www.python.org/>

⁵https://github.com/pinbo/gene_manual_annotation

⁶<http://202.194.139.32/searchtools/>

TABLE 1 | Summary statistics for heading date, plant height, thousand grain weight, and spike length for the two parents and the RIL population in 2016–2018.

Trait	Year	Parent		RIL population				
		<i>eh1</i>	LX987	Average	Min	Max	CV%	h^2
Heading date (d)	2016	202.4	208.2	205.8	202.0	210.0	0.8	0.8
	2017	198.1	206.7	202.7	197.0	210.0	1.4	
	2018	208.8	212.3	210.6	206.0	215.0	0.7	
Plant height (cm)	2016	67.9	86.8	75.2	48.8	100.0	14.4	0.82
	2017	76.4	86.7	82.2	63.7	106.0	9.5	
	2018	66.9	80.9	76.5	54.0	95.8	9.1	
Thousand grain weight (g)	2016	39.9	52.4	44.8	28.4	60.0	11.4	0.77
	2017	36.4	48.5	41.8	24.6	58.1	13.4	
	2018	33.3	44.9	38.5	24.8	53.6	13.0	
Spike length (cm)	2016	7.5	8.6	8.1	5.0	11.7	12.8	0.79
	2018	7.7	9.2	8.5	6.0	11.3	11.2	

Min and Max represent minimum and maximum of the corresponding traits among the RIL population. CV is coefficients of variation and h^2 is broad sense heritability.

²<http://www.cgmb.com.cn/>

³<http://polymarker.tgac.ac.uk/>

In addition, PH, TGW, and SL from 2016–2018 followed a normal distribution and strong transgressive segregation was observed in the RIL population (**Figure 1**). Analysis of variance of PH, TGW, and SL for the multiple environment trials in the RIL population indicated that these traits were affected by environmental conditions (**Supplementary Table 3**).

Analysis of the pairwise correlations between HD, PH, TGW, and SL suggested that TGW and SL were significantly positively correlated with PH while SL was significantly negatively correlated with HD (**Table 2**). However, the correlations between SL and PH, SL and HD were weak (**Table 2**).

Genetic Map Construction

Among the 400 RILs, 207 lines were randomly selected for genotyping using a Wheat55K SNP Array with 53,063 tags selected from the Wheat660K SNP Array (Ren et al., 2018). Since PHR SNPs are recommended for polyploid species and have the highest reliability, only PHR SNP probes were kept. SNPs with the same genotype in both parents were removed. Finally, 6505 SNP markers were obtained for genetic map construction (**Table 3**). These markers were divided into 1097 unique loci with the number distributed on each chromosome ranging from 10 to 96 (**Table 3**). The genetic

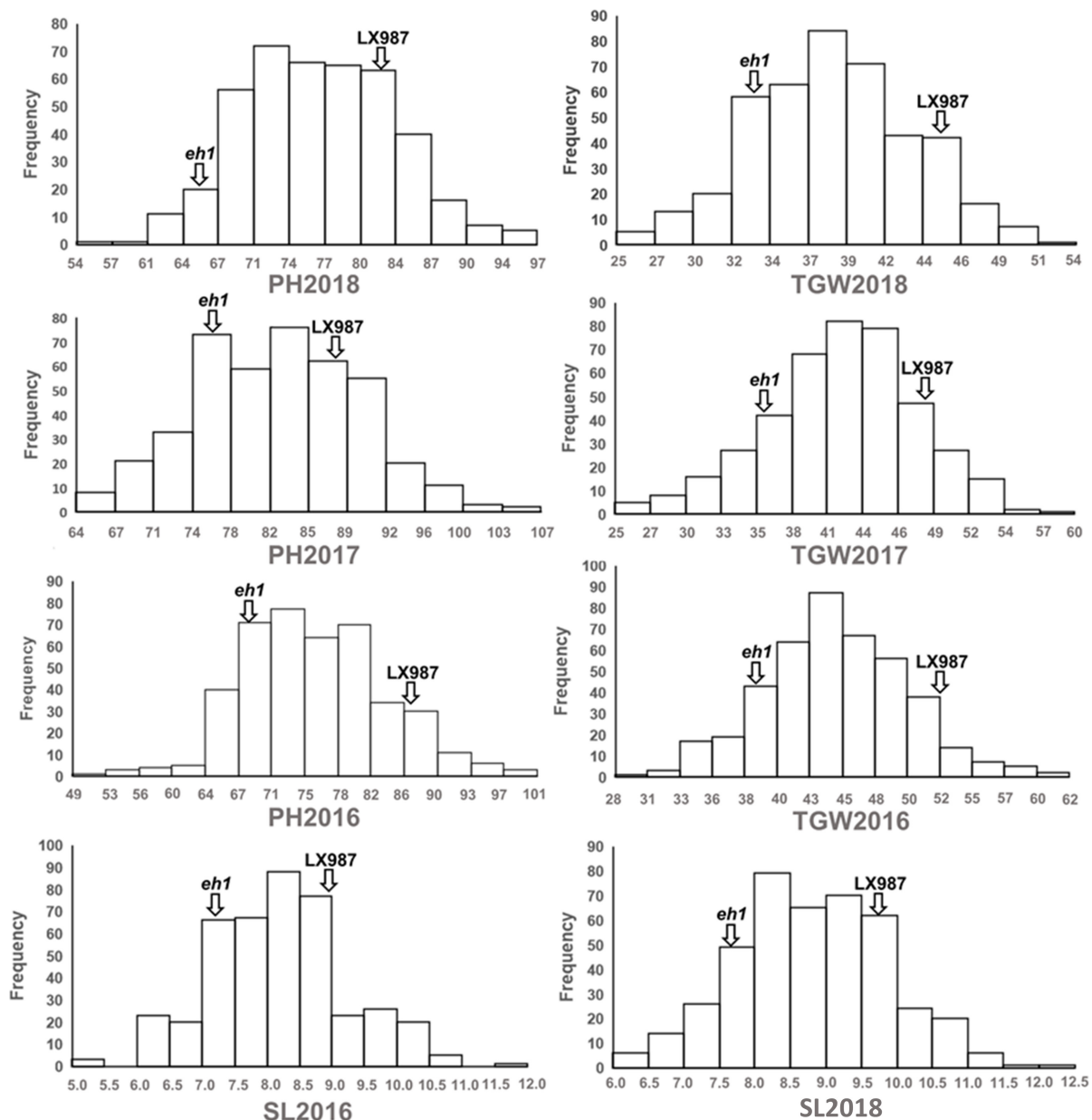


FIGURE 1 | Phenotypic of distribution of plant height (PH), thousand grain weight (TGW), and spike length (SL) in the RIL population from 2016 to 2018. Phenotypic values of the two parents were marked by vertical arrows.

TABLE 2 | Correlation coefficient analyses among heading date (HD), plant height (PH), thousand grain weight (TGW), and spike length (SL) in the RIL population.

Traits	HD	PH	TGW	SL
HD	1			
PH	0.0356 ^{ns}	1		
TGW	0.0799 ^{ns}	0.5570***	1	
SL	-0.2311***	0.1568**	0.0124 ^{ns}	1

Three lines from each year were used, and the correlations are the average values of the three years. ^{ns} means no significant correlation detected by statistical analysis. *** and ** indicate significance at $p \leq 0.001$ and $p \leq 0.01$, respectively.

TABLE 3 | Distribution of markers on 21 chromosomes in the constructed genetic map.

Chromosomes	No. of markers with poly high resolution	No. of unique loci	Length (cM)	cM/loci
1A	293	53	185.5	3.5
1B	164	55	110.3	2.0
1D	38	18	141.3	7.9
2A	1106	56	91.7	1.6
2B	328	80	123.5	1.5
2D	37	20	192.8	9.6
3A	392	84	243.7	2.9
3B	432	68	104.1	1.5
3D	36	14	145.2	10.4
4A	842	89	222.9	2.5
4B	277	47	86.0	1.8
4D	16	10	131.7	13.2
5A	535	95	215.9	2.3
5B	543	96	156.6	1.6
5D	50	22	264.7	12.0
6A	139	35	168.6	4.8
6B	345	49	119.2	2.4
6D	24	15	159.6	10.6
7A	458	84	254.0	3.0
7B	426	90	157.8	1.8
7D	24	17	221.3	13.0
Total	6505	1097	3496.1	5.2

map spanned 3496.1 cM in length with an average density of 5.2 cM/locus (Table 3).

QTL Mapping Analysis

A total of 37 QTLs for HD, PH, and TGW from 2016–2018, and SL from 2016 and 2018, were identified by QTL mapping analysis (Table 4 and Figure 2). These QTLs with LOD values ranging from 2.8 to 38.9 were distributed on 15 chromosomes and explained 3.0–48.8% of the phenotypic variation (Table 4 and Figure 2). There were 9, 7, 12, and 9 QTLs detected for HD, PH, TGW, and SL, respectively (Table 4 and Figure 2).

QTLs for HD were detected on chromosomes 1B (2), 2B (2), 3A (2), 4A (1), 5B (1), and 6B (1) (Table 4 and Figure 2). Notably, *qHD5B* and *qHD6B* were detected in all 3 years. *qHD5B* explained

18.4–48.8% of the phenotypic variation while *qHD6B* accounted for 3.3–11.6% of the phenotypic variation (Table 4). *qHD3A.1* was located close to *qHD3A.2*, and *qHD2B.1* was located close to *qHD2B.2*. *qHD2B.1* was detected in 2018 and explained 3.6% of the phenotypic variation, and *qHD2B.2* was detected in 2017 and explained 3.1% of the phenotypic variation (Table 4). For all of the QTLs except *qHD6B* the allele increasing HD was contributed by LX987 (Table 4).

For PH, 7 QTLs were identified on chromosomes 2A (1), 4A (1), 4B (2), and 6B (3) (Table 4 and Figure 2). *qPH4B.1* was identified in 2017 and 2018 (Table 4 and Figure 2), with LOD scores of 15.0 and 17.9 and explaining 31.5% and 34.4% of the phenotypic variation, respectively (Table 4). Three co-located QTLs were identified on chromosome 6B from 2016–2018, which explained 4.4%–8.8% of the variation in PH (Table 4 and Figure 2). For all of the QTLs except *qPH4A* the allele increasing PH was contributed by LX987 (Table 4).

For TGW, 12 QTLs were detected on chromosomes 3A (1), 3B (3), 3D (1), 4B (2), 5D (1), 6A (2), 7A (1), and 7D (1) (Table 4 and Figure 2). *qTGW4B.1* was detected in 2017 and 2018, explaining 9.6% and 22.5% of the variation in TGW, respectively (Table 4). *qTGW4B.2*, which was located close to *qTGW4B.1*, was detected in 2016 and explained 20.6% of the phenotypic variation (Table 4). In addition, the QTLs *qTGW3B.1* and *qTGW3B.2* were located close to each other and explained 5.5% and 13.1% of phenotypic variation in 2018 and 2017, respectively (Table 4). For all of the QTLs except *qTGW7A* and *qTGW3A*, the allele increasing TGW was contributed by LX987 (Table 4).

For SL, 9 QTLs were identified on chromosomes 3A (1), 4A (1), 5B (3), 6A (1), 6B (1), and 7D (2). *qSL6A* was detected in 2016 and 2018, explaining 9.5% and 22.0% of phenotypic variation, respectively (Table 4). *qSL5B.2* and *qSL5B.3* showed high contributions to phenotypic variation, 11.6% and 9.6%, respectively (Table 4). For *qSL7D.1*, *qSL5B.2*, *qSL5B.3*, and *qSL7D.2*, the negative alleles were contributed by LX987 while for the other QTLs the allele increasing SL was contributed by LX987 (Table 4).

Four QTL clusters were identified on chromosomes 3A, 4B, 5B, and 6B (Table 5). For the QTL cluster on chromosome 3A, *qHD3A.1* and *qHD3A.2* were co-localized with *qTGW3A* and *qSL3A* in a region ranging from 70.28 cM to 88.01 cM. On chromosome 4B, *qPH4B.1* for PH was clustered with two QTLs for TGW, with the alleles from LX987 increasing PH and TGW. For the QTL cluster on chromosome 5B, *qHD5B*, which was detected in all 3 years, was clustered with *qSL5B.3* (Tables 4, 5); however, the positive alleles for these QTLs were derived from opposite parents (Table 4). Three QTLs for PH on chromosome 6B were clustered with *qHD6B* and *qSL6B*, with the alleles from LX987 increasing PH and SL (Tables 4, 5).

To evaluate the QTL $\times$ environment interactions, Multi-Environment Traits (MET) analysis was employed by using QTL IciMapping v4.1 (Li et al., 2015). Similarly, 33 QTLs were identified by MET analysis (Supplementary Table 4). Among them, 10 QTLs showed significant interactions with environment, including the major QTLs *qHD5B*, *qPH4B.1*, and *qTGW4B.1*.

TABLE 4 | QTLs for heading date (HD), plant height (PH), thousand grain weight (TGW), and spike length (SL) in 2016, 2017, and 2018 identified by IciMapping 4.1.

Trait and year	QTL	Position	Left marker	Right marker	LOD ^a	PVE (%) ^b	Add ^c
HD2018	<i>qHD2B.1</i>	59	AX-110960849	AX-110972149	2.8	3.6	0.3
	<i>qHD3A.1</i>	81	AX-109580196	AX-110551014	4.1	4.6	0.4
	<i>qHD5B</i>	88	AX-111538681	AX-109870696	13.6	18.4	0.7
	<i>qHD6B</i>	117	AX-86183685	AX-110689596	9.0	11.6	−0.6
HD2017	<i>qHD2B.2</i>	68	AX-110983186	AX-111716247	3.3	3.1	0.5
	<i>qHD3A.2</i>	78	AX-109103063	AX-108814864	4.3	3.5	0.6
	<i>qHD4A</i>	149	AX-108917261	AX-108763710	4.1	3.2	0.6
	<i>qHD5B</i>	88	AX-111538681	AX-109870696	38.9	48.8	2.2
HD2016	<i>qHD6B</i>	114	AX-86183685	AX-110689596	3.9	3.3	−0.6
	<i>qHD1B.1</i>	2	AX-110448009	AX-109817665	3.3	3.3	0.3
	<i>qHD1B.2</i>	110	AX-109622448	AX-109490479	5.9	6.0	0.5
	<i>qHD5B</i>	88	AX-111538681	AX-109870696	32.3	39.8	1.3
PH2018	<i>qHD6B</i>	116	AX-86183685	AX-110689596	6.8	6.4	−0.5
	<i>qPH2A</i>	65	AX-108805248	AX-110671547	3.8	6.2	1.9
	<i>qPH4A</i>	133	AX-111567358	AX-109398960	4.5	6.0	−1.8
	<i>qPH4B.1</i>	51	AX-111542943	AX-86175614	17.9	34.4	4.1
PH2017	<i>qPH6B.1</i>	118	AX-86183685	AX-110689596	3.1	4.4	1.5
	<i>qPH4B.1</i>	51	AX-111542943	AX-86175614	15.0	31.5	4.0
	<i>qPH6B.2</i>	102	AX-110122533	AX-111092305	3.0	5.6	1.7
	<i>qPH6B.3</i>	111	AX-111092305	AX-86183685	4.0	8.8	3.6
PH2016	<i>qPH4B.2</i>	36	AX-109458638	AX-111074167	3.7	8.3	3.4
	<i>qPH6B.3</i>	111	AX-111092305	AX-86183685	4.0	8.8	3.6
TGW2018	<i>qTGW3B.1</i>	34	AX-109536560	AX-111059512	3.9	5.5	1.1
	<i>qTGW3D</i>	51	AX-110234451	AX-94739884	4.9	8.5	1.4
	<i>qTGW4B.1</i>	50	AX-109637078	AX-111542943	15.2	22.5	2.3
	<i>qTGW6A.1</i>	109	AX-110937386	AX-111013769	3.6	5.5	1.1
TGW2017	<i>qTGW7A</i>	138	AX-108759584	AX-111680717	4.0	5.8	−1.2
	<i>qTGW7D</i>	0	AX-108815937	AX-111379517	4.3	6.2	1.2
	<i>qTGW3B.2</i>	37	AX-110971226	AX-110375013	5.4	13.1	2.1
	<i>qTGW4B.1</i>	50	AX-109637078	AX-111542943	4.5	9.6	1.9
TGW2016	<i>qTGW5D</i>	12	AX-111577847	AX-95658716	4.0	9.4	1.8
	<i>qTGW6A.2</i>	168	AX-109868276	AX-109392684	3.2	7.5	1.6
	<i>qTGW3A</i>	75	AX-89583101	AX-111611367	3.6	5.8	−1.2
	<i>qTGW3B.3</i>	24	AX-109418825	AX-108890155	5.5	9.5	1.5
SL2018	<i>qTGW4B.2</i>	52	AX-108892921	AX-108871853	12.0	20.6	2.4
	<i>qSL4A</i>	126	AX-111124943	AX-108829087	3.5	3.8	0.2
	<i>qSL6A</i>	29	AX-109505625	AX-111507391	18.4	22.0	0.5
	<i>qSL7D.1</i>	165	AX-110975128	AX-111014383	5.4	6.0	−0.2
SL2016	<i>qSL3A</i>	88	AX-109584650	AX-109983808	4.2	3.0	0.2
	<i>qSL5B.1</i>	11	AX-109329070	AX-110531191	7.1	5.4	0.3
	<i>qSL5B.2</i>	25	AX-109431199	AX-108985377	14.2	11.6	−0.4
	<i>qSL5B.3</i>	86	AX-108872409	AX-109581384	11.9	9.6	−0.4
	<i>qSL6A</i>	30	AX-109505625	AX-111507391	11.1	9.5	0.4
	<i>qSL6B</i>	110	AX-110122533	AX-111092305	12.4	10.1	0.4
	<i>qSL7D.2</i>	1	AX-108815937	AX-111379517	6.7	5.1	−0.3

The QTLs written in bold font were detected in more than 1 year. ^aLOD values of detected QTL. ^bPhenotypic variation explained by the corresponding QTL. ^cAdditive effect of the corresponding QTL. Positive values indicate the alleles from LX987 increase the trait values, and negative values indicate the alleles from eh1 increase the corresponding trait values.

QTL Validation by Mapping With Molecular Markers

The QTLs on chromosomes 3A (*qHD3A*), 4B (*qPH4B.1* and *qTGW4B.1*), 5B (*qHD5B*), and 6A (*qSL6A*) were stably detected in different years. We selected these QTLs for validation using

KASP markers developed based on RNA-seq data (Li et al., 2020). In a recent study we reported that the *VRN-B1* gene located on chromosome 5B around the *qHD5B* region was responsible for HD variation in the RIL population (Li et al., 2020). For the validation of *qHD3A*, we successfully developed seven KASP

markers around or in the 55K SNP array-mapped region, and delimited the QTL to a genetic interval of 1.29 cM between markers 3A128b and 3A16, spanning approximately 2.5 Mb (Figure 3A). The LOD scores of this QTL were 5.7 and 7.5,

explaining 6.0% and 8.0% of the variation of HD, in 2017 and 2018, respectively (Table 6). For the QTLs on chromosome 4B, nine KASP markers were successfully developed, and the QTLs for PH were narrowed down to a genetic interval of 1.11 cM

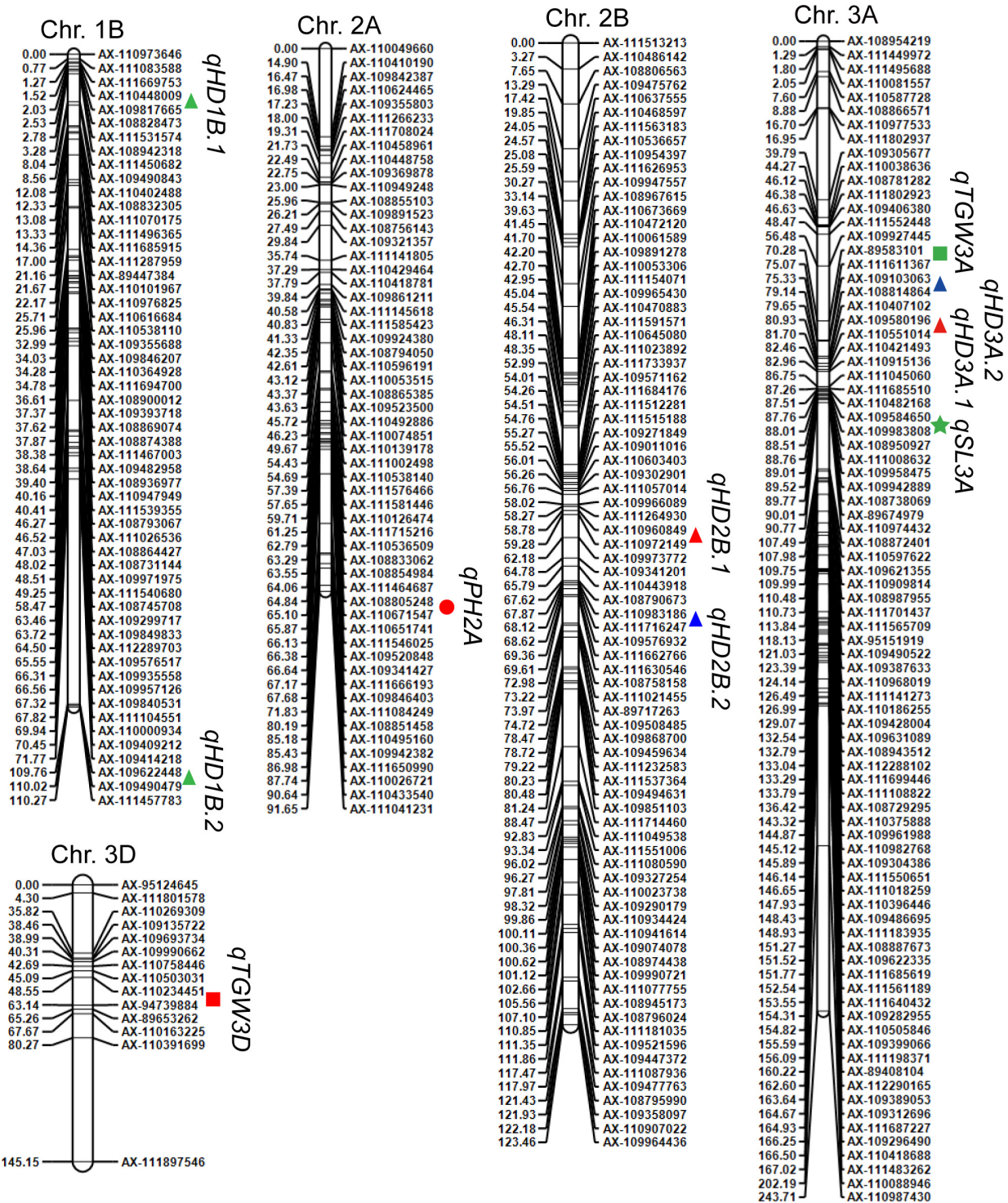


FIGURE 2 | Continued

flanked by markers 4B271b and 4B288b, corresponding to an approximately 0.8 Mb physical region (Figure 3B). This QTL was detected in 2017 and 2018, with LOD scores of 22.1 and 19.8, and explaining 22.8% and 20.5% of the variation in PH, respectively

(Table 6). Consistent with this, the QTL for TGW identified in 2016 and 2018 was mapped between markers 4B271b and 4B288b with LOD scores of 11.1 and 16.8 and explaining 12.2% and 17.8% of the variation in TGW, respectively (Figure 3B and

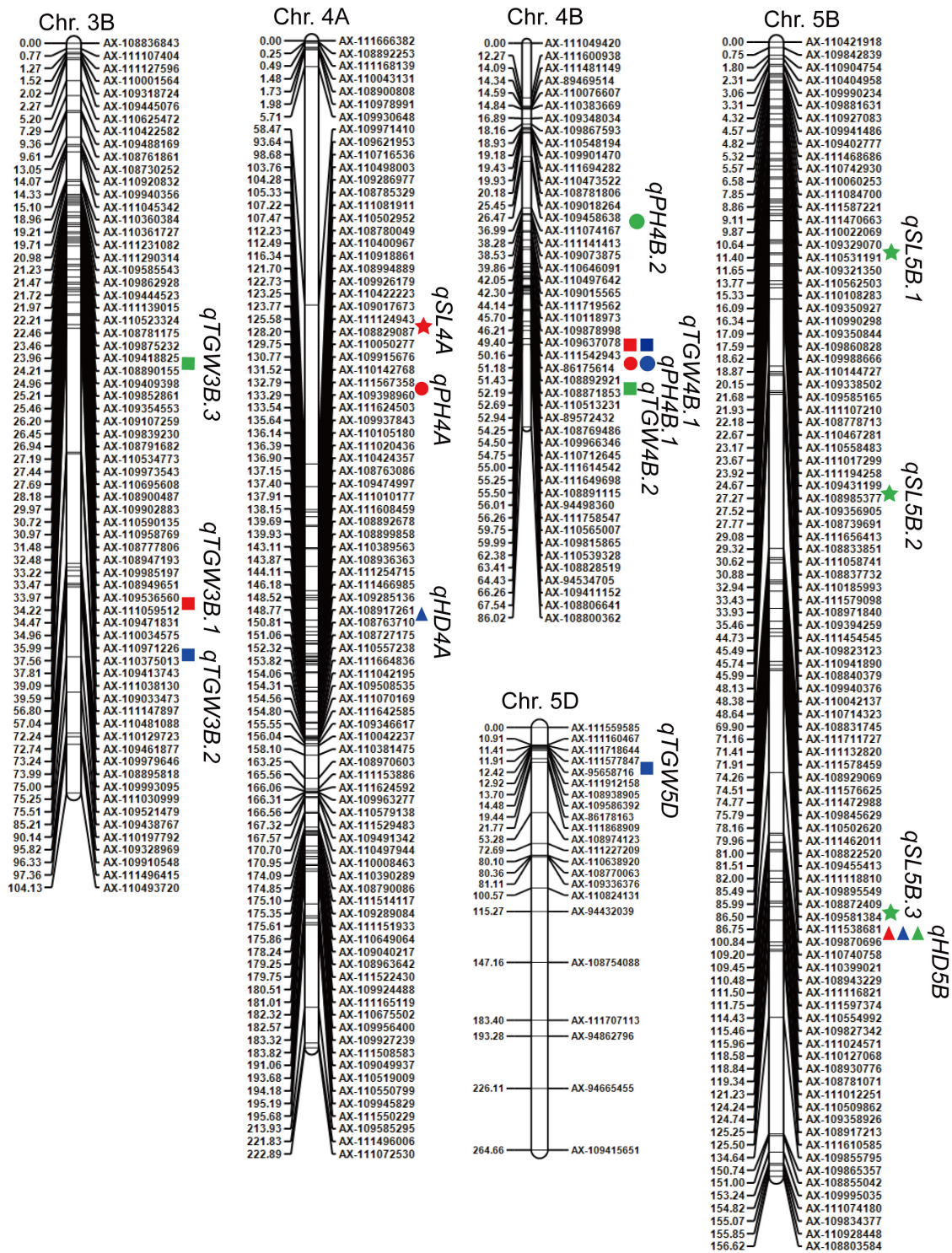


FIGURE 2 | Continued

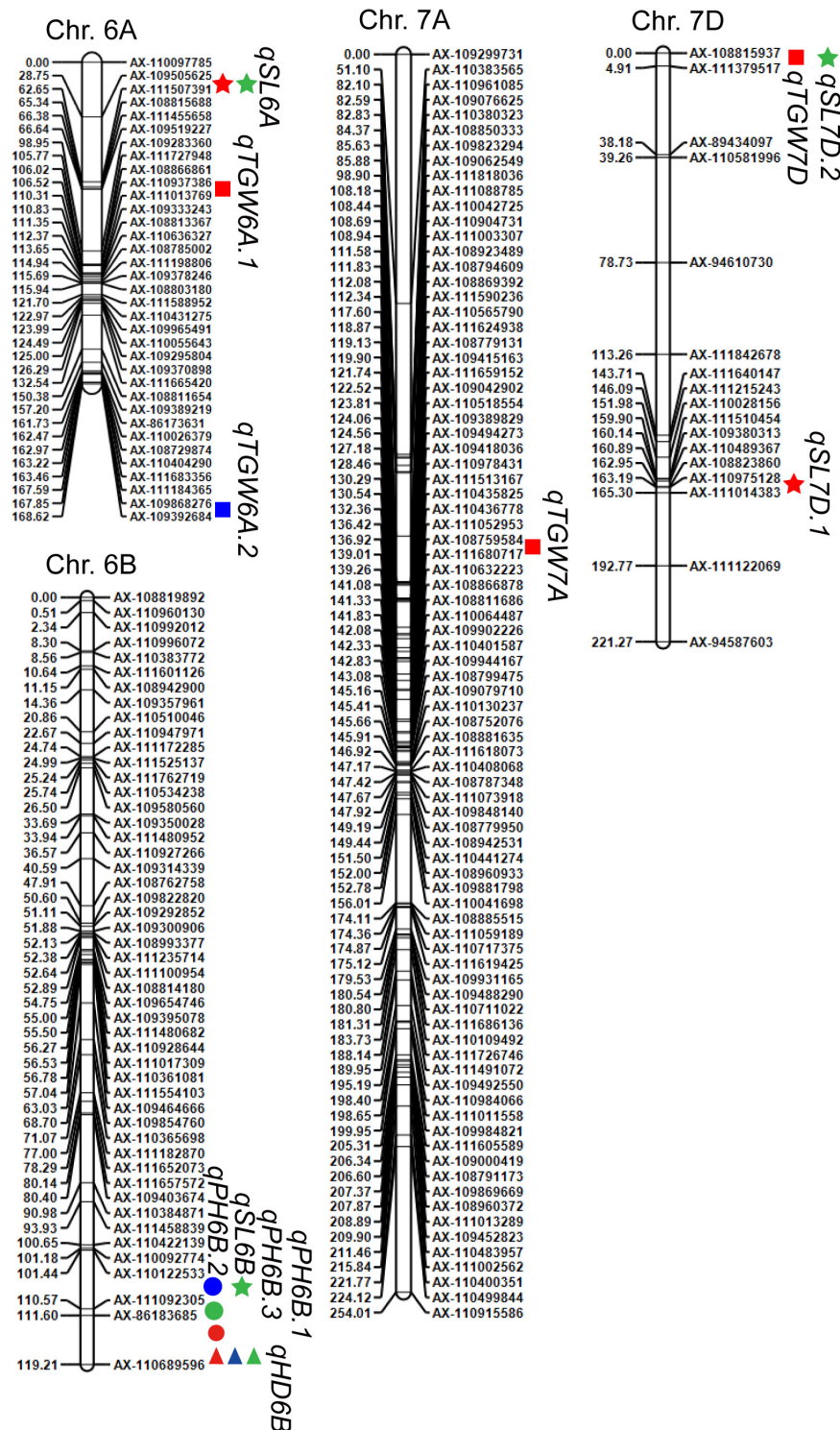


FIGURE 2 | Chromosomal locations of the identified QTLs for HD, PH, TGW, and SL in 2016, 2017, and 2018. Triangles, circles, squares, and stars represent HD, PH, TGW, and SL, respectively. The colors green, blue, and red indicate data from 2016, 2017, and 2018, respectively.

Table 6). For the validation of *qSL6A*, we successfully developed 10 KASP markers for genetic mapping. In 2016 and 2018, a major QTL for SL with LOD scores of 12.6 and 23.5 and

explaining 13.1% and 22.5% of phenotypic variation, respectively, was detected between markers 6A51 and 6A419 within a genetic interval of 31.8 cM (**Figure 3C** and **Table 6**). Due to the large

TABLE 5 | QTL clusters affecting two or more traits. QTLs from each year located within 10 cM and affected two or more traits were identified as a QTL cluster.

Chromosome	QTL	Marker interval	Position (cM)
3A	<i>qHD3A.1</i> , <i>qHD3A.2</i> , <i>qTGW3A</i> , <i>qSL3A</i>	AX-89583101–AX-109983808	70.28–88.01
4B	<i>qPH4B.1</i> , <i>qTGW4B.1</i> , <i>qTGW4B.2</i>	AX-109637078–AX-108871853	49.40–52.19
5B	<i>qHD5B</i> , <i>qSL5B.3</i>	AX-108872409–AX-109870696	85.99–100.84
6B	<i>qHD6B</i> , <i>qPH6B.1</i> , <i>qPH6B.2</i> , <i>qPH6B.3</i> , <i>qSL6B</i>	AX-110122533–AX-110689596	101.44–119.21

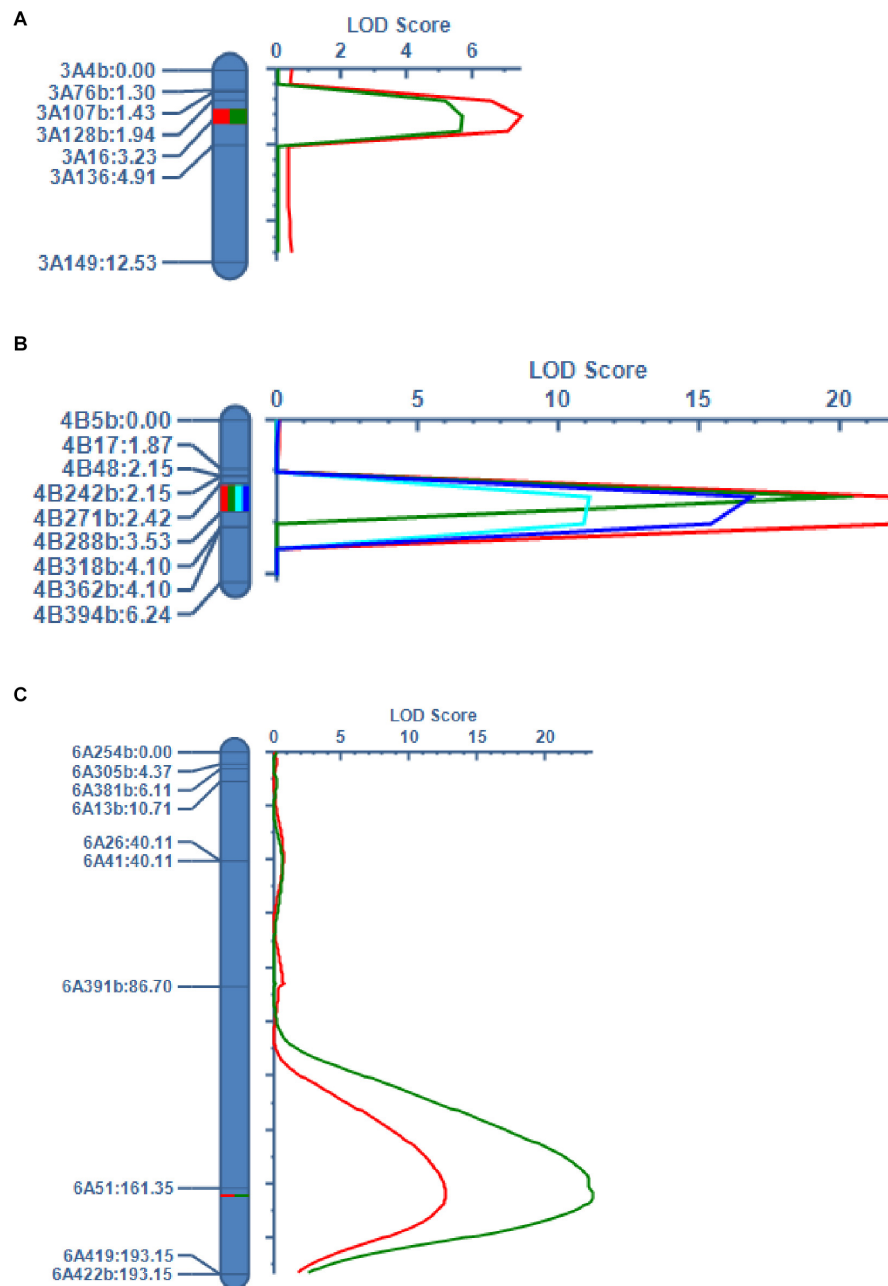
**FIGURE 3** | QTL validation using KASP markers developed based on RNA-seq data. **(A)** QTLs for HD on chromosome 3A. The green and red curves represent LOD scores from 2017 and 2018, respectively. **(B)** QTLs for PH and TGW on chromosome 4B. The red and green curves represent LOD scores for PH from 2017 and 2018, respectively. The light blue and blue curves represent LOD scores for TGW from 2016 and 2018, respectively. **(C)** QTLs for SL on chromosome 6A. The red and green curves represent LOD scores from 2016 and 2018, respectively. The marker name and genetic position of each marker are indicated on the left side of each chromosome.

TABLE 6 | Details of the genetic map for QTL validation generated using KASP markers.

Trait name	Chromosome	Position	Left marker	Right marker	LOD	PVE (%)	Add
HD2018	3A	3	3A128b	3A16	7.5	8.0	0.4
HD2017	3A	3	3A128b	3A16	5.7	6.0	0.8
PH2017	4B	3	4B271b	4B288b	22.1	22.8	3.7
PH2018	4B	3	4B271b	4B288b	19.8	20.5	3.2
TGW2016	4B	3	4B271b	4B288b	11.1	12.2	1.8
TGW2018	4B	3	4B271b	4B288b	16.8	17.8	2.1
SL2016	6A	164	6A51	6A419	12.6	13.1	0.4
SL2018	6A	164	6A51	6A419	23.5	22.5	0.5

The 400 RILs were used for QTL validation.

region for this QTL, we did not conduct further analysis of the candidate genes.

Gene Annotations and Effects of SNPs in the Validated QTL Regions on Chromosome 4B and 3A

Since *qPH4B.1* and *qTGW4B.1* were delimited to a physical interval of 0.8 Mb between markers 4B271b and 4B288b by genetic mapping with KASP markers, we analyzed gene models and annotations in this region according to the Chinese Spring (CS) reference genome v1.0 (International Wheat Genome Sequencing Consortium, 2018). In this region, seven high-confidence genes were annotated. Based on BLASTP searches for rice and *Arabidopsis* homologous genes⁷, these genes are predicted to encode 40S ribosomal protein S27 (TraesCS4B01G280800), Beta-galactosidase (TraesCS4B01G280900), a Histidine-containing phosphotransfer protein (TraesCS4B01G281000), 60S ribosomal protein L5 (TraesCS4B01G281100), Protein PAM68 (TraesCS4B01G281200), Tribbles homolog 3 (TraesCS4B01G281300), and a Ubiquitin carboxyl-terminal hydrolase family protein (TraesCS4B01G281400) (Supplementary Table 5). We also analyzed sequence variation between LX987 and *eh1* in this region based on RNA-seq data. A total of 18 SNPs with each parent homozygous for different alleles were identified (Supplementary Table 6). Analysis of SNP effects suggested that three SNPs were missense mutations. One SNP in TraesCS4B02G281200 located at the 189th position caused a change in the amino acid Leu in LX987 to Trp in *eh1* and was predicted to have the largest effect on gene function. Multiple alignment of amino acid sequences of protein PAM68 from grasses indicated that this region is conserved among *Brachypodium distachyon*, *Sorghum bicolor*, *Zea mays* and rice (Supplementary Figure 2).

For the HD QTL on chromosome 3A between markers 3A128b and 3A16, we found that 38 high-confidence genes were annotated in the mapped interval (Supplementary Table 7). In this region, nine homozygous SNPs with genotypes differing between the two parent lines were found based on RNA-seq

data. One SNP in TraesCS3A01G086400, which encodes an R3H domain-containing protein, that caused a change from Ser to Pro at the 267th position had the largest effect on gene function (Supplementary Table 8).

DISCUSSION

QTL Mapping Using the WHEAT55K SNP Array

SNP arrays are a powerful and effective approach for QTL mapping (Rasheed et al., 2017). The tags of the Wheat55K Array (Affymetrix® Axiom® Wheat55) were carefully selected from the Wheat660K Array, and all tags were uniformly distributed on 21 chromosomes. Therefore, the 55K Array is suitable for genotyping in QTL studies (Ren et al., 2018). The Wheat55K SNP Array has been utilized for QTL mapping of productive tiller number (Liu et al., 2018b), temporal expression of tiller number (Ren et al., 2018), and leaf rust and stripe rust resistance (Huang et al., 2019; Zhang et al., 2019) in wheat. In this study, we used the Wheat55K SNP Array to genotype 207 RILs and constructed a genetic map containing 6,505 PHR SNP markers (Table 3). PHR SNPs are of high quality and possess better cluster resolution than other SNPs (Marrano et al., 2019), which improves the accuracy of genotyping. The genetic map spanned 3496.1 cM across the 21 chromosomes, which is similar to the total length of genetic maps for 199 wheat RILs constructed by Liu et al. (2018b) and 186 RILs constructed by Huang et al. (2019). We detected a total of 37 QTLs for HD, PH, TGW, and SL by mapping using the 55K SNP array (Table 4 and Figure 2). Among these QTLs, those on chromosomes 3A (*qHD3A*), 4B (*qPH4B.1* and *qTGW4B.1*), 5B (*qHD5B*) (Li et al., 2020), and 6A (*qSL6A*) that were stably detected in different years, were validated using KASP markers (Figure 3). High LOD values ranging from 5.7 to 23.5 were observed for the QTLs that were validated with KASP markers (Table 6), indicating that the QTLs detected using the 55K SNP array data are reliable.

Comparison of the Mapped QTLs With Those Identified in Previous Studies

A total of nine QTLs for HD were mapped on chromosomes 1B, 2B, 3A, 4A, 5B, and 6B (Table 4 and Figure 2). Consistent with these findings, in our previous study we also identified QTLs for HD on chromosomes 2B, 3A, and 5B using BSA of the same RIL population (Li et al., 2020). In addition, *qHD1B.1* and *qHD4A* were mapped to genetic regions similar to those reported by Zhao et al. (2019). *qHD2B.2* was mapped to a genetic position similar to that of HD QTLs reported by Hu et al. (2020) and Li et al. (2018). The analysis of the physical positions of the flanking markers in the wheat reference genome indicated that *qHD2B.2* is probably the *Ppd-B1* gene. The two adjacent QTLs *qHD3A.1* and *qHD3A.2* are located at a genetic position similar to that of an HD QTL reported by Li et al. (2018). The QTL *qHD5B*, which was stably detected in different years (Table 4 and Figure 2), is located around gene *VRN-B1*, and our previous results suggested that the *VRN-B1* gene is responsible for HD variation in this

⁷<http://202.194.139.32/searchtools/>

RIL population (Li et al., 2020). In addition, the stably detected QTL *qHD6B* was found at a position similar to that of HD QTLs reported by Perez-Lara et al. (2016) and Li et al. (2018).

Regarding PH, we identified two and three QTLs on chromosomes 4B and 6B, respectively (Table 4 and Figure 2). Consistent with these results, previous studies also reported several PH QTLs on chromosomes 4B and 6B (Gao et al., 2016; Li et al., 2018; Jahani et al., 2019). *qPH4B.2*, *qPH6B.1*, *qPH6B.2*, and *qPH6B.3* were mapped to similar genetic positions on chromosomes 4B and 6B as PH QTLs reported by Gao et al. (2015); Li et al. (2018), and Hu et al. (2020). BLAST searches of flanking markers for *qPH4B.2* against the wheat reference genome indicated that this region harbors the reported *RhtB1b* gene (Peng et al., 1999). However, using KASP markers for *Rht1* (Rasheed et al., 2016) we found that both of the parent lines, *eh1* and LX987, had the same *RhtB1b* genotype (Supplementary Figure 1), indicating that the detection of *qPH4B.2* is not due to *RhtB1b*. It is possible that there is other variation in the *Rht1* gene that causes differences in PH between the parent lines. *qPH2A* is located at a similar genetic position as a QTL reported by Li et al. (2018), and the *qPH4A* region overlaps with the physical region reported by Chen et al. (2020). It has been reported that the semi-dominant dwarfing gene *Rht-NM9* is located in a region from 178.9 Mb to 187.2 Mb on 2AS (Lu et al., 2015). We performed a BLAST search using the flanking markers for *qPH2A* and found that this QTL is located in a 143–148 Mb interval according to the CS reference genome. Therefore, we speculate that *qPH2A* does not harbor the *Rht-NM9* gene.

The stably detected QTL *qTGW4B.1* co-localized with *qTGW4B.2* (Table 4 and Figure 2). These QTLs are located within 49.4–52.19 cM. Guan et al. (2018) reported stable QTLs for TGW located within 22.3–95.8 cM on chromosome 4B. *qTGW3A* is located at a genetic position similar to that reported in Cui et al. (2014a). *qTGW3D* is located at positions similar to those reported in Cui et al. (2014b) and Gao et al. (2015). The QTLs *qTGW3B.1*, *qTGW3B.2*, *qTGW3B.3*, *qTGW5D*, and *qTGW7A* are located at positions similar to those reported by Li et al. (2018), and *qTGW6A.1* is located close to a stable yield and TGW QTL reported by Simmonds et al. (2014). In addition, the QTLs *qTGW7A* and *qTGW7D* are located at similar genetic positions as those reported by Cui et al. (2014a) and Guan et al. (2018), respectively.

qSL4A is located at a genetic position similar to that reported by Cui et al. (2012b) and Gao et al. (2015). *qSL5B.2* is located at a position similar to that in Cui et al. (2012b). In addition, *qSL6B* is located at a genetic position similar to that reported by Li et al. (2018) and Hu et al. (2020). To the best of our knowledge, the stably detected QTL *qSL6A* with a LOD value ranging from 11.1 to 18.4 is likely to be a new QTL (Table 4 and Figure 2).

Pleiotropic QTLs for HD, PH, TGW, and SL

Among the QTLs for HD, PH, TGW, and SL detected in this study, four regions controlled two or more of these traits (Table 5). In addition to *Rht1*, a previous study identified a “QTL-hotspot” region for yield-related traits on chromosome 4B (Guan et al., 2018). This is consistent with the QTL cluster

detected in our study (Table 5). Consistent with the positive correlation between PH and TGW (Table 2 and Supplementary Table 9), the superior alleles of the co-localized QTLs *qPH4B.1*, *qTGW4B.1*, and *qTGW4B.2* were derived from the same parent line (Table 4). A QTL cluster for HD, PH, and SL that mapped to the interval 101.44–119.21 cM on chromosome 6B (Table 5) is likely the same or similar to a QTL cluster for yield-related traits reported by Li et al. (2018). *qHD6B* co-localized with *qSL6B*, with favorable alleles derived from opposite parents (Tables 4, 5). This is consistent with the negative correlation between HD and SL (Table 2).

Candidate Genes Affecting PH, TGW, and HD

Using KASP markers, we delimited the QTL regions for PH and TGW on chromosome 4B to a 0.8 Mb physical region (Figure 3B and Table 6). A recent study identified a QTL cluster for TGW linked to *Rht-B1* on chromosome 4B using near-isogenic lines (Guan et al., 2020). This region includes the physical interval identified in our study. According to gene annotation and analysis of the effects of SNPs in the mapped region (Supplementary Table 6), a mutation in TraesCS4B02G281200 encoding a PAM68 protein showed the largest effect on gene function. The PAM68 protein is essential for efficient D1 biogenesis and photosystem II assembly in *Arabidopsis* (Armbruster et al., 2010, 2013). Split-ubiquitin assays suggested that the C terminus of *Arabidopsis* PAM68 is required for interaction with the PSII core proteins D1 and CP43 (Armbruster et al., 2010). The variation in the PAM68 protein between LX987 and *eh1* is located at the C terminus, and this region is conserved in grasses (Supplementary Figure 2). This indicates that the mutation in PAM68 may affect gene function. Photosynthesis plays an important role in yield improvement (Zhu et al., 2010). It has been reported that mutation of the photosystem 1-F subunit (*OsPS1-F*) results in reduction of PH and grain yield in rice (Ramamoorthy et al., 2018). Taken together, our results and previous findings suggest that PAM68 is a candidate gene for the PH and TGW QTLs. We found genes TraesCS4B01G281000 and TraesCS4B01G281300 in the QTL region were not expressed in the RNA-seq data, which may be due to that the RNA-seq data was collected from spikes in the HD (Li et al., 2020). Therefore, we could not exclude these two genes as candidate genes from the sequences of RNA and the predicted effects of SNP. However, TraesCS4B01G281000 and TraesCS4B01G281300 encode Histidine-containing phosphotransfer protein and Tribbles homolog 3, respectively, which have not been reported for involving in PH and TGW. In this respect, the possibility for these two candidate genes is low.

We also confirmed and narrowed down the QTL region for HD on chromosome 3A to a 2.5 Mb interval (Figure 3A and Table 6). Analysis of SNP effects suggested that a mutation in TraesCS3A01G086400 has a large effect on gene function (Supplementary Table 8), suggesting that this gene may affect HD in the RIL population. TraesCS3A01G086400 encodes an R3H domain-containing protein, which functions in binding polynucleotides, including DNA, RNA, and single-stranded DNA (Grishin, 1998). Studies on R3H-containing proteins in maize

(Saleh et al., 2006) and *Arabidopsis* (Wang et al., 2019) have suggested that these proteins are involved in stress responses. Whether the R3H domain-containing protein contributes to HD variation need to be further studied.

CONCLUSION

We identified 37 QTLs for HD, PH, TGW, and SL in a RIL population using the Wheat55K SNP Array, and validated the stably detected QTLs on chromosome 3A, 4B, and 6A using KASP markers. The QTLs on chromosomes 4B and 3A were delimited to a physical interval of 0.8 Mb and 2.5 Mb, respectively. Moreover, the candidate genes affecting PH, TGW, and HD were predicted based on gene annotation and analysis of SNP effects. The linked KASP markers developed in this study will facilitate breeding for yield improvement in wheat.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

LL conceived the project and revised the manuscript. HX and YL conducted most of the experiments. HG constructed the RIL population and assisted in collection of the phenotypic data. YX and JG performed genotype analyses. LZ, SZ, and YD participated in field trials. HX wrote the first draft of the manuscript. All authors have read and approved the final manuscript.

FUNDING

This work was financially supported by the National Natural Science Foundation of China (grant numbers 31801346 and 11775304), the National Key Research and Development Program of China (grant numbers 2020YFE0202300), the China Agriculture Research System (grant number CARS-03), and

Fundamental Research Funds for Central Non-Profit of Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (grant number Y2020YJ09).

ACKNOWLEDGMENTS

We thank Dr. Junli Zhang (Department of Plant Sciences, University of California, Davis, United States) for assisting in SNP effects analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.628478/full#supplementary-material>

Supplementary Figure 1 | Genotypes of *eh1* and LX987 obtained using the *Rht1* KASP markers.

Supplementary Figure 2 | Multiple sequence alignment of PAM68 in *Brachypodium distachyon*, barley, rice, sorghum, maize, and wheat. The labeling with red asterisk indicates the position of variation between the two parent lines.

Supplementary Table 1 | Physical positions of left and right markers of identified QTLs.

Supplementary Table 2 | The primers used for KASP assays.

Supplementary Table 3 | Variance analysis of PH, TGW, and SL in the 400 RIL. *** indicates significance at $p \leq 0.001$.

Supplementary Table 4 | MET analysis of QTLs by environment interactions. QTL with a LOD (AbyE) >2.5 was significant interaction with environment.

Supplementary Table 5 | High-confidence genes located between markers 4B271b and 4B288b.

Supplementary Table 6 | SNPs identified using RNA-seq data for LX987 and *eh1* in the mapped region of chromosome 4B.

Supplementary Table 7 | High-confidence genes located between markers 3A128b and 3A16.

Supplementary Table 8 | SNPs identified using RNA-seq data for LX987 and *eh1* in the mapped region of chromosome 3A.

Supplementary Table 9 | Correlation coefficient analyses of HD, PH, TGW, and SL among different years.

REFERENCES

- Alvarez, M. A., Tranquilli, G., Lewis, S., Kippes, N., and Dubcovsky, J. (2016). Genetic and physical mapping of the earliness *per se* locus Eps-A (m) 1 in *Triticum monococcum* identifies EARLY FLOWERING 3 (ELF3) as a candidate gene. *Funct. Integr. Genom.* 16, 365–382. doi: 10.1007/s10142-016-0490-3
- Araki, E., Miura, H., and Sawada, S. (1999). Identification of genetic loci affecting amylose content and agronomic traits on chromosome 4A of wheat. *Theor. Appl. Genet.* 98, 977–984. doi: 10.1007/s001220051158
- Armbruster, U., Rühle, T., Kreller, R., Strotbek, C., Zühlke, J., Tadini, L., et al. (2013). The photosynthesis affected mutant68-like protein evolved from a PSII assembly factor to mediate assembly of the chloroplast NAD(P)H dehydrogenase complex in *Arabidopsis*. *Plant Cell* 25, 3926–3943. doi: 10.1105/tpc.113.114785
- Armbruster, U., Zühlke, J., Rengstl, B., Kreller, R., Makarenko, E., Rühle, T., et al. (2010). The *Arabidopsis* thylakoid protein PAM68 is required for efficient D1 biogenesis and photosystem II assembly. *Plant Cell* 22, 3439–3460. doi: 10.1105/tpc.110.077453
- Asplund, L., Leino, M., and Hagenblad, J. (2012). Allelic variation at the *Rht8* locus in a 19th century wheat collection. *Sci. World J.* 2012:385610. doi: 10.1100/2012/385610
- Beales, J., Turner, A., GriYths, S., Snape, J. W., and Laurie, D. A. (2007). A Pseudo-Response Regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 115, 721–733.
- Bednarek, J., Boulaflous, A., Girousse, C., Ravel, C., Tassy, C., Barret, P., et al. (2012). Down-regulation of the TaGW2 gene by RNA interference results in decreased grain size and weight in wheat. *J. Exp. Bot.* 63, 5945–5955.

- Börner, A., Schumann, E., Fürste, A., Cöster, H., Leithold, B., Röder, M., et al. (2002). Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 105, 921–936. doi: 10.1007/s00122-002-0994-1
- Cao, P., Liang, X., Zhao, H., Feng, B., Xu, E., Wang, L., et al. (2019). Identification of the quantitative trait loci controlling spike-related traits in hexaploid wheat (*Triticum aestivum* L.). *Planta* 250, 1967–1981. doi: 10.1007/s00425-019-03278-0
- Chai, L., Chen, Z., Bian, R., Zhai, H., Cheng, X., Peng, H., et al. (2019). Dissection of two quantitative trait loci with pleiotropic effects on plant height and spike length linked in coupling phase on the short arm of chromosome 2D of common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 132, 1815–1831. doi: 10.1007/s00122-019-03318-z
- Chen, S., Gao, R., Wang, H., Wen, M., Xiao, J., Bian, N., et al. (2015). Characterization of a novel reduced height gene (Rht23) regulating panicle morphology and plant architecture in bread wheat. *Euphytica* 203, 583–594. doi: 10.1007/s10681-014-1275-1
- Chen, Z., Cheng, X., Chai, L., Wang, Z., Bian, R., Li, J., et al. (2020). Dissection of genetic factors underlying grain size and fine mapping of QTgw.cau-7D in common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 133, 149–162. doi: 10.1007/s00122-019-03447-5
- Cui, F., Ding, A., Li, J., Zhao, C., Wang, L., Wang, X., et al. (2012a). QTL detection of seven spike-related traits and their genetic correlations in wheat using two related RIL populations. *Euphytica* 186, 177–192. doi: 10.1007/s10681-011-0550-7
- Cui, F., Ding, A. M., Li, J., Zhao, C. H., Wang, L., Wang, X. Q., et al. (2012b). QTL detection of seven spike-related traits and their genetic correlations in wheat using two related RIL populations. *Euphytica* 186, 177–192.
- Cui, F., Zhao, C., Ding, A., Li, J., Wang, L., Li, X., et al. (2014a). Construction of an integrative linkage map and QTL mapping of grain yield-related traits using three related wheat RIL populations. *Theor. Appl. Genet.* 127, 659–675. doi: 10.1007/s00122-013-2249-8
- Cui, F., Zhao, C. H., Ding, A. M., Li, J., Wang, L., Li, X. F., et al. (2014b). Construction of an integrative linkage map and QTL mapping of grain yield-related traits using three related wheat RIL populations. *Theor. Appl. Genet.* 127, 659–675.
- Cuthbert, J. L., Somers, D. J., Brûlé-Babel, A. L., Brown, P. D., and Crow, G. H. (2008). Molecular mapping of quantitative trait loci for yield and yield components in spring wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 117, 595–608. doi: 10.1007/s00122-008-0804-5
- Deng, S., Wu, X., Wu, Y., Zhou, R., Wang, H., Jia, J., et al. (2011). Characterization and precise mapping of a QTL increasing spike number with pleiotropic effects in wheat. *Theor. Appl. Genet.* 122, 281–289. doi: 10.1007/s00122-010-1443-1
- Ellis, M. H., Rebetzke, G. J., Azanza, F., Richards, R. A., and Spielmeyer, W. (2005). Molecular mapping of gibberellin-responsive dwarfing genes in bread wheat. *Theor. Appl. Genet.* 111, 423–430. doi: 10.1007/s00122-005-2008-6
- Flintham, J. E., Börner, A., Worland, A. J., and Gale, M. D. (1997). Optimizing wheat grain yield: effects of Rht (gibberellin-insensitive) dwarfing genes. *J. Agr. Sci.* 128, 11–25.
- Gao, F., Wen, W., Liu, J., Rasheed, A., Yin, G., Xia, X., et al. (2015). Genome-wide linkage mapping of QTL for yield components, plant height and yield-related physiological traits in the Chinese wheat cross Zhou 8425B/Chinese Spring. *Front. Plant Sci.* 6:1099. doi: 10.3389/fpls.2015.01099
- Gao, F. M., Liu, J. D., Yang, L., Wu, X. X., Xiao, Y. G., Xia, X. C., et al. (2016). Genome-wide linkage mapping of QTL for physiological traits in a Chinese wheat population using the 90K SNP array. *Euphytica* 209, 789–804.
- Grant, N. P., Mohan, A., Sandhu, D., and Gill, K. S. (2018). Inheritance and genetic mapping of the reduced height (Rht18) gene in wheat. *Plants* 7:58. doi: 10.3390/plants7030058
- Grishin, N. V. (1998). The R3H motif: a domain that binds single-stranded nucleic acids. *Trends Biochem. Sci.* 23, 329–330.
- Guan, P., Lu, L., Jia, L., Kabir, M. R., Zhang, J., Lan, T., et al. (2018). Global QTL analysis identifies genomic regions on chromosomes 4A and 4B harboring stable loci for yield-related traits across different environments in wheat (*Triticum aestivum* L.). *Front. Plant Sci.* 9:529. doi: 10.3389/fpls.2018.00529
- Guan, P., Shen, X., Mu, Q., Wang, Y., Wang, X., Chen, Y., et al. (2020). Dissection and validation of a QTL cluster linked to Rht-B1 locus controlling grain weight in common wheat (*Triticum aestivum* L.) using near-isogenic lines. *Theor. Appl. Genet.* 133, 2639–2653. doi: 10.1007/s00122-020-03622-z
- Haq, M. A., Martinek, P., Watanabe, N., and Kuboyama, T. (2011). Genetic mapping of gibberellin acid-sensitive genes for semi-dwarfism in durum wheat. *Cereal Res. Commun.* 39, 171–178.
- Hu, J., Wang, X., Zhang, G., Jiang, P., Chen, W., Hao, Y., et al. (2020). QTL mapping for yield-related traits in wheat based on four RIL populations. *Theor. Appl. Genet.* 133, 917–933. doi: 10.1007/s00122-019-03515-w
- Huang, S., Wu, J., Wang, X., Mu, J., Xu, Z., Zeng, Q., et al. (2019). Utilization of the genomewide wheat 55K SNP array for genetic analysis of stripe rust resistance in common wheat Line P9936. *Phytopathology* 109, 819–827. doi: 10.1094/PHYTO-10-18-0388-R
- International Wheat Genome Sequencing Consortium. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191
- Jahani, M., Mohammadi-Nejad, G., Nakhoda, B., and Rieseberg, L. H. (2019). Genetic dissection of epistatic and QTL by environment interaction effects in three bread wheat genetic backgrounds for yield-related traits under saline conditions. *Euphytica* 215:103. doi: 10.1007/s10681-019-2426-1
- Jaiswal, V., Gahlaut, V., Mathur, S., Agarwal, P., Khandelwal, M. K., Khurana, J. P., et al. (2015). Identification of novel SNP in promoter sequence of TaGW2-6A associated with grain weight and other agronomic traits in wheat (*Triticum aestivum* L.). *PLoS One* 10:e0129400. doi: 10.1371/journal.pone.0129400
- Kawaura, K., Takaku, M., Imai, T., and Ogihara, Y. (2009). Molecular analysis of the Q gene controlling spike morphology with TILLING lines of common wheat. *Genes Genet. Syst.* 84, 476–476.
- Kippes, N., Debernardi, J. M., Vasquez-Gross, H. A., Akpinar, B. A., Budak, H., Kato, K., et al. (2015). Identification of the VERNALIZATION 4 gene reveals the origin of spring growth habit in ancient wheats from south Asia. *Proc. Natl. Acad. Sci. U.S.A.* 112, E5401–E5410.
- Kiseleva, A. A., and Salina, E. A. (2018). Genetic regulation of common wheat heading time. *Russ. J. Genet.* 54, 375–388. doi: 10.1134/S1022795418030067
- Kumar, A., Mantovani, E. E., Seetan, R., Soltani, A., Echeverry-Solarte, M., Jain, S., et al. (2016). Dissection of genetic factors underlying wheat kernel shape and size in an elite x non-adapted cross using a high density SNP linkage map. *Plant Genome* 9:81. doi: 10.3835/plantgenome2015.09.0081
- Kumar, N., Kulwal, P. L., Gaur, A., Tyagi, A. K., Khurana, J. P., Khurana, P., et al. (2006). QTL analysis for grain weight in common wheat. *Euphytica* 151, 135–144.
- Li, F., Wen, W., He, Z., Liu, J., Jin, H., Cao, S., et al. (2018). Genome-wide linkage mapping of yield-related traits in three Chinese bread wheat populations using high-density SNP markers. *Theor. Appl. Genet.* 131, 1903–1924. doi: 10.1007/s00122-018-3122-6
- Li, S., Wang, J., and Zhang, L. (2015). Inclusive composite interval mapping of QTL by environment interactions in biparental populations. *PLoS One* 10:e0132414.
- Li, W. L., Nelson, J. C., Chu, C. Y., Shi, L. H., Huang, S. H., Liu, D. J., et al. (2002). Chromosomal locations and genetic relationships of tiller and spike characters in wheat. *Euphytica* 125, 357–366.
- Li, Y., Xiong, H., Guo, H., Zhou, C., Xie, Y., Zhao, L., et al. (2020). Identification of the vernalization gene VRN-B1 responsible for heading date variation by QTL mapping using a RIL population in wheat. *BMC Plant Biol.* 20:331. doi: 10.1186/s12870-020-02539-5
- Liu, J., Xu, Z. B., Fan, X. L., Zhou, Q., Cao, J., Wang, F., et al. (2018a). A genome-wide association study of wheat spike related traits in China. *Front. Plant Sci.* 9:1584. doi: 10.3389/fpls.2018.01584
- Liu, J. J., Luo, W., Qin, N. N., Ding, P. Y., Zhang, H., Yang, C. C., et al. (2018b). A 55K SNP array-based genetic map and its utilization in QTL mapping for productive tiller number in common wheat. *Theor. Appl. Genet.* 131, 2439–2450.
- Lou, X., Li, X., Li, A., Pu, M., Shoaib, M., Liu, D., et al. (2016). The 160 bp insertion in the promoter of Rht-B1i plays a vital role in increasing wheat height. *Front. Plant Sci.* 7:307. doi: 10.3389/fpls.2016.00307

- Lu, Y., Xing, L., Xing, S., Hu, P., Cui, C., Zhang, M., et al. (2015). Characterization of a putative new semi-dominant reduced height gene, *Rht*_NM9, in wheat (*Triticum aestivum* L.). *J. Genet. Genom.* 42, 685–698. doi: 10.1016/j.jgg.2015.08.007
- Ma, J., Zhang, H., Li, S., Zou, Y., Li, T., Liu, J., et al. (2019). Identification of quantitative trait loci for kernel traits in a wheat cultivar Chuannong16. *BMC Genet.* 20:77. doi: 10.1186/s12863-019-0782-4
- Marrano, A., Martínez-García, P. J., Bianco, L., Sideli, G. M., Di Pierro, E. A., Leslie, C. A., et al. (2019). A new genomic tool for walnut (*Juglans regia* L.): development and validation of the high-density Axiom™ J. regia 700K SNP genotyping array. *Plant Biotechnol. J.* 17, 1027–1036. doi: 10.1111/pbi.13034
- Marza, F., Bai, G. H., Carver, B. F., and Zhou, W. C. (2006). Quantitative trait loci for yield and related traits in the wheat population Ning7840 x Clark. *Theor. Appl. Genet.* 112, 688–698. doi: 10.1007/s00122-005-0172-3
- Mayer, K. F. X., Rogers, J., Doležal, J., Pozniak, C., Eversole, K., Feuillet, C., et al. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345:6194. doi: 10.1126/science.1251788
- Milec, Z., Valarik, M., Bartos, J., and Safar, J. (2014). Can a late bloomer become an early bird? Tools for flowering time adjustment. *Biotechnol. Adv.* 32, 200–214. doi: 10.1016/j.biotechadv.2013.09.008
- Mir, R. R., Kumar, N., Jaiswal, V., Girdharwal, N., Prasad, M., Balyan, H. S., et al. (2012). Genetic dissection of grain weight in bread wheat through quantitative trait locus interval and association mapping. *Mol. Breed.* 29, 963–972.
- Mo, Y., Vanzetti, L. S., Hale, I., Spagnolo, E. J., Guidobaldi, F., Al-Oboudi, J., et al. (2018). Identification and characterization of *Rht25*, a locus on chromosome arm 6AS affecting wheat plant height, heading time, and spike development. *Theor. Appl. Genet.* 131, 2021–2035.
- Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., et al. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* 400, 256–261. doi: 10.1038/22307
- Perez-Lara, E., Semagn, K., Chen, H., Iqbal, M., N'Diaye, A., Kamran, A., et al. (2016). QTLs associated with agronomic traits in the Cutler x AC barrie spring wheat mapping population using single nucleotide polymorphic markers. *PLoS One* 11:e0160623. doi: 10.1371/journal.pone.0160623
- Ramamoorthy, R., Vishal, B., Ramachandran, S., and Kumar, P. P. (2018). The OsPSI-F gene regulates growth and development in rice by modulating photosynthetic electron transport rate. *Plant Cell Rep.* 37, 377–385. doi: 10.1007/s00299-017-2235-8
- Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R. K., et al. (2017). Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. *Mol. Plant* 10, 1047–1064. doi: 10.1016/j.molp.2017.06.008
- Rasheed, A., Wen, W., Gao, F., Zhai, S., Jin, H., Liu, J., et al. (2016). Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. *Theor. Appl. Genet.* 129, 1843–1860. doi: 10.1007/s00122-016-2743-x
- Rasheed, A., and Xia, X. (2019). From markers to genome-based breeding in wheat. *Theor. Appl. Genet.* 132, 767–784. doi: 10.1007/s00122-019-03286-4
- Ren, T., Hu, Y., Tang, Y., Li, C., Yan, B., Ren, Z., et al. (2018). Utilization of a Wheat55K SNP array for mapping of major QTL for temporal expression of the tiller number. *Front. Plant Sci.* 9:333. doi: 10.3389/fpls.2018.00333
- Saleh, A., Lumberras, V., Lopez, C., Dominguez-Puigianer, E., Kizis, D., Pages, M., et al. (2006). Maize DBF1-interactor protein 1 containing an R3H domain is a potential regulator of DBF1 activity in stress responses. *Plant J.* 46, 747–757. doi: 10.1111/j.1365-3113.2006.02742.x
- Simmonds, J., Scott, P., Brinton, J., Mestre, T. C., Bush, M., del Blanco, A., et al. (2016). A splice acceptor site mutation in TaGW2-A1 increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. *Theor. Appl. Genet.* 129, 1099–1112.
- Simmonds, J., Scott, P., Leverington-Waite, M., Turner, A. S., Brinton, J., Korzun, V., et al. (2014). Identification and independent validation of a stable yield and thousand grain weight QTL on chromosome 6A of hexaploid wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 14:191. doi: 10.1186/s12870-014-0191-9
- Snape, J. W., Butterworth, K., Whitechurch, E., and Worland, A. J. (2001). Waiting for fine times: genetics of flowering time in wheat. *Euphytica* 119, 185–190. doi: 10.1023/A:1017594422176
- Su, Z. Q., Hao, C. Y., Wang, L. F., Dong, Y. C., and Zhang, X. Y. (2011). Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 122, 211–223.
- Tian, X. L., Wen, W. E., Xie, L., Fu, L. P., Xu, D. G., Fu, C., et al. (2017). Molecular mapping of reduced plant height gene *Rht24* in bread wheat. *Front. Plant Sci.* 8:1379. doi: 10.3389/fpls.2017.01379
- Tshikunde, N. M., Mashilo, J., Shimelis, H., and Odindo, A. (2019). Agronomic and physiological traits, and associated quantitative trait loci (QTL) affecting yield response in wheat (*Triticum aestivum* L.): a review. *Front. Plant Sci.* 10:1428. doi: 10.3389/fpls.2019.01428
- Varshney, R. K., Prasad, M., Roy, J. K., Kumar, N., Harjit, S., Dhaliwal, H. S., et al. (2000). Identification of eight chromosomes and a microsatellite marker on 1AS associated with QTL for grain weight in bread wheat. *Theor. Appl. Genet.* 100, 1290–1294. doi: 10.1007/s001220051437
- Vikhe, P., Patil, R., Chavan, A., Oak, M., and Tamhankar, S. (2017). Mapping gibberellin-sensitive dwarfing locus *Rht18* in durum wheat and development of SSR and SNP markers for selection in breeding. *Mol. Breed.* 37, 1–10.
- Wang, H. Y., Liu, C., Ren, Y. C., Wu, M. H., Wu, Z. W., Chen, Y., et al. (2019). An RNA-binding protein MUG13.4 interacts with AtAGO2 to modulate salinity tolerance in *Arabidopsis*. *Plant Sci.* 288:110218. doi: 10.1016/j.plantsci.2019.110218
- Wolde, G. M., Trautewig, C., Mascher, M., and Schnurbusch, T. (2019). Genetic insights into morphometric inflorescence traits of wheat. *Theor. Appl. Genet.* 132, 1661–1676. doi: 10.1007/s00122-019-03305-4
- Wurschum, T., Langer, S. M., Longin, C. F. H., Tucker, M. R., and Leiser, W. L. (2017). A modern green revolution gene for reduced height in wheat. *Plant J.* 92, 892–903. doi: 10.1111/tpj.13726
- Xie, L., Zhang, Y., Wang, K., Luo, X., Xu, D., Tian, X., et al. (2019). TaVrt2, an SVP-like gene, cooperates with TaVrn1 to regulate vernalization-induced flowering in wheat. *New Phytol.* 2019:16339. doi: 10.1111/nph.16339
- Xu, D., Wen, W., Fu, L., Li, F., Li, J., Xie, L., et al. (2019). Genetic dissection of a major QTL for kernel weight spanning the *Rht-B1* locus in bread wheat. *Theor. Appl. Genet.* 132, 3191–3200. doi: 10.1007/s00122-019-03418-w
- Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., et al. (2006). The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 19581–19586.
- Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., and Dubcovsky, J. (2003). Positional cloning of the wheat vernalization gene *VRN1*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6263–6268.
- Yan, L. L., Loukoianov, A., Blechl, A., Tranquilli, G., Ramakrishna, W., SanMiguel, P., et al. (2004). The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science* 303, 1640–1644.
- Yang, J., Zhou, Y., Wu, Q., Chen, Y., Zhang, P., Zhang, Y., et al. (2019). Molecular characterization of a novel TaGL3-5A allele and its association with grain length in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 132, 1799–1814. doi: 10.1007/s00122-019-03316-1
- Yang, L., Zhao, D., Meng, Z., Xu, K., Yan, J., Xia, X., et al. (2020). QTL mapping for grain yield-related traits in bread wheat via SNP-based selective genotyping. *Theor. Appl. Genet.* 133, 857–872. doi: 10.1007/s00122-019-03511-0
- Yao, H., Xie, Q., Xue, S., Luo, J., Lu, J., Kong, Z., et al. (2019). HL2 on chromosome 7D of wheat (*Triticum aestivum* L.) regulates both head length and spikelet number. *Theor. Appl. Genet.* 132, 1789–1797. doi: 10.1007/s00122-019-03315-2
- Zhai, H. J., Feng, Z. Y., Du, X. F., Song, Y. E., Liu, X. Y., Qi, Z. Q., et al. (2018). A novel allele of TaGW2-A1 is located in a finely mapped QTL that increases grain weight but decreases grain number in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 131, 539–553.
- Zhang, P., Li, X., Gebrewahid, T. W., Liu, H., Xia, X., He, Z., et al. (2019). QTL mapping of adult-plant resistance to leaf and stripe rust in wheat cross SW 8588/Thatcher using the wheat 55K SNP array. *Plant Dis.* 103, 3041–3049. doi: 10.1094/pdis-02-19-0380-re
- Zhang, Y., Li, D., Zhang, D. B., Zhao, X. G., Cao, X. M., Dong, L. L., et al. (2018). Analysis of the functions of TaGW2 homoeologs in wheat grain weight and protein content traits. *Plant J.* 94, 857–866.

- Zhao, C. H., Sun, H., Liu, C., Yang, G. M., Liu, X. J., Wang, Y. P., et al. (2019). Detection of quantitative trait loci for wheat (*Triticum aestivum* L.) heading and flowering date. *J. Agr. Sci.* 157, 20–30. doi: 10.1017/S0021859619000200
- Zhu, X. G., Long, S. P., and Ort, D. R. (2010). Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.* 61, 235–261. doi: 10.1146/annurev-arplant-042809-112206
- Zikhali, M., Leverington-Waite, M., Fish, L., Simmonds, J., Orford, S., Wingen, L. U., et al. (2014). Validation of a 1DL earliness *per se* (eps) flowering QTL in bread wheat (*Triticum aestivum*). *Mol. Breed.* 34, 1023–1033. doi: 10.1007/s11032-014-0094-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Xiong, Li, Guo, Xie, Zhao, Gu, Zhao, Ding and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genome-Wide Identification and Low Temperature Responsive Pattern of Actin Depolymerizing Factor (ADF) Gene Family in Wheat (*Triticum aestivum* L.)

Ke Xu^{1†}, Yong Zhao^{1**}, Sihang Zhao¹, Haodong Liu¹, Weiwei Wang^{1,2}, Shuhua Zhang¹ and Xueju Yang^{1*}

OPEN ACCESS

Edited by:

Xianchun Xia,
Institute of Crop Sciences, Chinese
Academy of Agricultural Sciences,
China

Reviewed by:

Mingming Xin,
China Agricultural University, China
Xiaohui Li,
Capital Normal University, China
Yuming Wei,
Sichuan Agricultural University, China

*Correspondence:

Yong Zhao
zhaoyong_0423@163.com
Xueju Yang
shmyxj@126.com

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 09 November 2020

Accepted: 05 February 2021

Published: 24 February 2021

Citation:

Xu K, Zhao Y, Zhao S, Liu H,
Wang W, Zhang S and Yang X (2021)
Genome-Wide Identification and Low
Temperature Responsive Pattern
of Actin Depolymerizing Factor (ADF)
Gene Family in Wheat (*Triticum*
aestivum L.).
Front. Plant Sci. 12:618984.
doi: 10.3389/fpls.2021.618984

¹ State Key Laboratory of North China Crop Improvement and Regulation, Hebei Agricultural University, Baoding, China,
² Cangzhou Academy of Agriculture and Forestry Sciences, Cangzhou, China

The actin depolymerizing factor (ADF) gene family, which is conserved in eukaryotes, is important for plant development, growth, and stress responses. Cold stress restricts wheat growth, development, and distribution. However, genome-wide identification and functional analysis of the ADF family in wheat is limited. Further, because of the promising role of ADF genes in cold response, there is need for an understanding of the function of this family on wheat under cold stress. In this study, 25 ADF genes (*TaADFs*) were identified in the wheat genome and they are distributed on 15 chromosomes. The *TaADF* gene structures, duplication events, encoded conserved motifs, and *cis*-acting elements were investigated. Expression profiles derived from RNA-seq data and real-time quantitative PCR analysis revealed the tissue- and temporal-specific *TaADF* expression patterns. In addition, the expression levels of *TaADF13/16/17/18/20/21/22* were significantly affected by cold acclimation or freezing conditions. Overexpression of *TaADF16* increased the freezing tolerance of transgenic *Arabidopsis*, possibly because of enhanced ROS scavenging and changes to the osmotic regulation in cells. The expression levels of seven cold-responsive genes were up-regulated in the transgenic *Arabidopsis* plants, regardless of whether the plants were exposed to low temperature. These findings provide fundamental information about the wheat ADF genes and may help to elucidate the regulatory effects of the encoded proteins on plant development and responses to low-temperature stress.

Keywords: actin depolymerizing factor, wheat (*Triticum aestivum* L.), low temperature, genome-wide identification, transgenic *Arabidopsis*

INTRODUCTION

Actin depolymerizing factor (ADF), which is a conserved protein family with a low molecular weight (15–22 kDa) (Staiger et al., 1997) in eukaryotic cells, is crucial for regulating the reorganization and rearrangement of the actin cytoskeleton (Maciver and Hussey, 2002). This protein was first identified in chick embryo brains (Bamburg et al., 1980), but it has since been

detected in various plants (Gungabissoon et al., 1998; Smertenko et al., 1998; Dong et al., 2001; Wang et al., 2009; Fu et al., 2014). As a member of the actin-binding protein family, ADF remodels the actin cytoskeleton via the transition between F-actin and G-actin states (Du et al., 2016). The actin cytoskeleton influences cellular architecture as well as diverse processes, including the expression of cell polarity, cell expansion and division, intracellular transport, pathogen perception, abiotic stress responses, and signal transduction (Fowler and Quatrano, 1997; Drobak et al., 2004; Porter and Day, 2016; Allard et al., 2020).

Because of the various actin cytoskeleton functions affecting plant development and responses to stimuli, ADF proteins in higher plants are important for various cellular activities. For example, ADFs are involved in pollen development and pollen tube growth in *Arabidopsis* (Daher and Geitmann, 2012; Zheng et al., 2013), maize (Lopez et al., 1996), and tobacco (Chen et al., 2002). Previous studies revealed that a lack of *AtADF9* results in early flowering (Burgos-Rivera et al., 2008) and knockout of *AtADF4* increases the length of hypocotyls and epidermal cells (Henty et al., 2011). In cotton, *GhADF1* is involved in fiber elongation and secondary cell wall formation (Wang et al., 2009). Additionally, the expression patterns of ADF genes appear to be tissue-specific. Earlier research indicated that *OsADF2/4/5* are expressed in roots, stems, blades, sheaths, spikelets and seeds persistently, whereas *OsADF9* expression is specific to spikelets at the heading stage in rice (Huang et al., 2012). The *GhADF6/8* genes are preferentially expressed in petals, while *GhADF7* is highly expressed in anthers of cotton (Li et al., 2010). Moreover, ADF proteins reportedly have a vital role in responses to abiotic and biotic stresses in *Arabidopsis*, wheat, rice, and other species (Ouellet et al., 2001; Huang et al., 2012; Fu et al., 2014; Sengupta et al., 2019). The overexpression of *OsADF3* enhances the drought tolerance of transgenic *Arabidopsis* seedlings (Huang et al., 2012). The expression of wheat *TaADF4* is induced by heat, but down-regulated by low temperatures or salt stress (Zhang et al., 2017). In contrast, *TaADF3* expression is significantly up-regulated under cold conditions and water deficiency treatment, but is relatively un-affected by wounding or salt stress (Tang et al., 2016). Both *TaADF4* and *TaADF7* enhance the resistance of wheat plants to a *Puccinia striiformis* f. sp. *tritici* (*Pst*) infection, whereas *TaADF3* has the opposite effect (Fu et al., 2014; Tang et al., 2016; Zhang et al., 2017).

Low-temperature stress is a key factor influencing the growth, yield, and distribution of wheat plants (Song et al., 2017). An exposure to low but nonfreezing temperatures (i.e., cold acclimation) is crucial for the freezing tolerance of overwintering plants (Thomashow, 2001; Sung and Amasino, 2005). Although there are relatively few reports describing ADFs in wheat exposed to abiotic stress, two earlier investigations proved that ADF production is induced at low temperatures (Danyluk et al., 1996; Ouellet et al., 2001). During the cold acclimation process, the depolymerization of microtubules and actin filaments increases the fluidity of the plasma membrane (Danyluk et al., 1996). However, the molecular mechanism regulating the depolymerization of microtubules and actin filaments under cold conditions remains unknown. Thus,

clarifying the relationship between ADF proteins and freezing tolerance response is warranted.

The development and application of genome sequencing technology has led to the identification of ADF genes in the genomes of several plant species, including rice, maize, tomato, and *Arabidopsis* (Feng et al., 2006; Khatun et al., 2016; Huang et al., 2020). The availability of a sequenced 'Chinese Spring' genome has helped to facilitate the genome-wide analysis of gene families in wheat (Hu et al., 2018; International Wheat Genome Sequencing Consortium (IWGSC), 2014; Liu et al., 2019; Zhou et al., 2019). However, to the best of our knowledge, there are no reports regarding the genome-wide identification and characterization of wheat ADF genes or the wheat ADF gene expression profiles in various tissues and in response to low-temperature stress. In this study, we identified 25 wheat ADF genes using the 'Chinese Spring' genome sequences (IWGSC, RefSeq V1.1), after which we analyzed the ADF gene structures and the encoded conserved motifs, determined the genomic locations and duplication events of the ADF genes, and predicted the putative *cis*-acting elements. Additionally, ADFs in *Triticum dicoccoides*, *Hordeum vulgare*, *Triticum turgidum*, *Triticum urartu*, and *Aegilops tauschii*, were identified and used along with the wheat ADFs to construct a phylogenetic tree. To further investigate the function of ADF in wheat, we analyzed the expression profiles of these genes in different tissues, as well as in response to cold stress. Furthermore, we examined the effects of *TaADF16* overexpression on the freezing tolerance of *Arabidopsis*. The results of this study will enhance our understanding of the wheat ADF gene family and provide the basis for future investigations of ADFs in wheat.

MATERIALS AND METHODS

Identification of ADF Genes in Wheat

The genome and protein sequences data of wheat were downloaded from Ensembl Plants database¹. The actin-depolymerizing factor homology domain (ADF-H domain, with Pfam: PF00241) obtained from PFAM database² was employed as a query for Hidden Markov Model (HMM) search using HMMER3.0 with a pre-defined threshold of E value $\leq 1e^{-10}$. The results obtained were used to construct a wheat-specific ADF HMM profile by hmm-build program, and the second HMM search was used to remove the redundant sequences among the identified ADF proteins with an E value $\leq 1e^{-10}$. After manual corrections applied as needed, the NCBI-CDD web server³, SMART database⁴ and Pfam database (see foot note 2) were used to further confirm the ADF_H domain in the putative ADF protein sequences. The biochemical parameters of TaADF proteins, such as isoelectric points (pI), molecular weights (MW), instability

¹<http://plants.ensembl.org/index.html>

²<http://pfam.xfam.org/>

³<https://www.ncbi.nlm.nih.gov/cdd/>

⁴<http://smart.embl-heidelberg.de/>

index (II), aliphatic index (AI) and calculated grand average of hydropathy index (GRAVY) of the putative ADF proteins were calculated using the ExPASy online protParam tool⁵ (Artimo et al., 2012). The prediction of subcellular location of the identified TaADFs were performed using Plant-mPLOC⁶ (Chou and Shen, 2010). Alignment analysis of TaADF proteins was performed using MEGA 7.0, and visualized by Jalview v2.11.1.3 (Waterhouse et al., 2009).

Phylogenetic Tree Construction of Wheat and Other Eight Species

The phylogenetic tree was performed using the neighbor joining (NJ) method in MEGA 7.0, with 1,000 bootstrap replicates. Sequences of ADF proteins from select species were identified based on the corresponding genome (Supplementary Table 1) as described above. The accuracy of identified ADF was confirmed with Ensembl Plants¹ and Uniprot database⁷.

Gene Structure, Motif Analysis and *Cis*-Acting Elements of TaADF Gene Family

The exon/intron structures of TaADFs were constructed by gene structure display server (GSDS) program⁸ using the CDS and corresponding genomic sequences retrieved from the Ensembl plants database. Conserved motifs of TaADF discover were predicted using the Multiple Expectation Maximization for Motif Elicitation (MEME) 4.12.0⁹, with the following parameters: maximum number of 20 motifs and optimum motif widths of 6–100 residues. The 1500 bp upstream of the transcription start site (−1) of all identified TaADF transcripts was extracted as promoter to predict *cis*-acting elements using Plant CARE¹⁰.

Chromosomal Localization, Gene Duplication and Calculating Ka/Ks Values of TaADF

All the *TaADF* genes were mapped to wheat chromosomes based on physical location information from the database of wheat genome. The gene duplication in the wheat genome were analyzed with Multiple Collinearity Scan toolkit (MCScanX) (Wang et al., 2012) and visualized with Circos tool (Krzywinski et al., 2009). The calculation of *ka* and *ks* substitution of each duplicated *TaADF* genes were performed using KaKs_Calculator 2.0 (Wang et al., 2010). The syntenic maps between wheat and other species were constructed using the python version of MCScanX¹¹.

Expression Pattern of *TaADF* Genes in Different Tissues and Development Stages

The expression patterns of identified *TaADF* genes in different tissues and development stages were analyzed based on the publicly available wheat RNA-Seq datasets obtained from wheat eFP Browser¹² (Ramírez-González et al., 2018). The expression levels were summarized as transcripts per million (TPM), and a heatmap of *TaADF* tissues-specific expression were conducted with R packages (Pheatmap and Stats).

Plant Materials and Growth Conditions of Wheat

For expression analysis of *TaADFs* of different tissues, seeds of wheat ‘Chinese Spring’ were grown under 20°C with a 12/12 h photoperiod/dark in glass dish for 15 day (three-leaf stage). Roots and leaves were collected from five seedlings. For low temperature treatment, the wheat seedlings of ‘Jing 411’ were cultivated in the incubator at 20°C with a 12/12 h photoperiod/dark period for 15 days until three-leaf stage (TL), which was followed with different temperature treatments: 4°C for cold acclimation (CA), 20°C for un-cold acclimation (UCA). After 28 days, the CA and UCA seedlings were exposed to −5°C for 1 day (cold acclimation and freezing, CAF; un-cold acclimation and freezing, UCAF). All the samples were immediately frozen in liquid nitrogen and stored at −80°C. Three biological replications were performed.

RNA Extraction, RNA-Seq and RT-PCR Validation

The samples were subjected to total RNA extraction using a Trizol Reagent (Invitrogen, Carlsbad, CA, United States). The analysis of RNA-seq for low temperature treatment was based on our previous study (Zhao et al., 2019b), we re-analyzed the data based on reference genome of ‘Chinese Spring’ genome (IWGSC: RefSeq V1.1). Sampled crowns from TL, CA, UCA, CAF, and UCAF were used for RNA-seq analysis. Using the DESeq R package, the differently expressed genes (DEGs, |log2 (fold change)| > 1 and *p* < 0.05) were analyzed.

cDNA products were subjected to RT-PCR analysis, in which, *TaGAPDH* and *TaTEF1-α* were used as double internal reference genes for wheat. RT-PCR was performed with BCS Wiz SYBR Green RT-PCR Master Mix and the QuantStudio 5 Real-time PCR system (Applied Biosystems, Malham, MA, United States). The following amplification protocol was used: first step, 95°C for 30 s; second step, 40 cycles of 95°C for 5 s and 60°C for 30 s; final step, 95°C for 15 s, 60°C for 1 min, 95°C for 15 s, and 50°C for 30 s. The relative expression were calculated with $2^{-\Delta\Delta C_t}$ method. Three biological replications were performed (each biological replication for three technical). Specific primers used in this study are shown in Supplementary Table 2.

⁵<https://web.expasy.org/protParam/>

⁶<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>

⁷<https://www.uniprot.org/>

⁸<http://gsds.cbi.pku.edu.cn/>

⁹<http://meme-suite.org/>

¹⁰<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>

¹¹[https://github.com/tanghaibao/jcvi/wiki/MCScan-\(Python-version\)](https://github.com/tanghaibao/jcvi/wiki/MCScan-(Python-version))

¹²http://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi

Subcellular Localization of TaADFs

The coding sequence (CDS) of *TaADF11*, *TaADF14*, *TaADF15*, and *TaADF16* was cloned into the pBWA(V)HS-ccdb-GLosgfp vector containing the cauliflower mosaic virus 35S (CaMV 35S) promoter, respectively. Subsequently, the control plasmid and fusion plasmids were transiently expressed in *A. thaliana* protoplasts. Then, the transformed protoplasts were incubated for 24 h at 22°C darkness. Finally, green fluorescent protein (GFP) fluorescence signals were observed using Nikon C2-ER confocal laser scanning confocal microscope. The specific primers containing the restriction site are shown in **Supplementary Table 2**.

Overexpression of *TaADF16* Genes in Arabidopsis and Freezing Tolerance Assay

The CDS of *TaADF16* (Gene id: TraesCS5A02G478500), was amplified by PCR with gene-specific primer (F: 5'-GGAGAGGACACGCTCGAGATGACTTTATCTCGCCGACATG-3' and R 5'-TTAAAGCAGGACTCTAGATTAGGTGGTGTAGTCCTTGAGGAT-3') and then cloned into the pART-CAM vector controlled by the CaMV 35S promoter. The 35S:*TaADF16* plasmid were transformed into *Agrobacterium tumefaciens* GV3101 and then transformed into *Arabidopsis*. Seeds of T0 transgenic plants were selected on MS medium containing 100 µg/ml kanamycin and further confirmed by PCR. T3 homozygous of *Arabidopsis* transgenic plants was used for freezing tolerance assay.

For freezing tolerance analysis in WT and OE lines, all the seeds were planted in plastic pots filled nutrient soil for three weeks, with 14/10 h photoperiod and temperature at 22°C. Three weeks old seedlings were transferred to -5°C for 4 h for freezing stress and ion leakage was determined. For recovery treatment, the plants after freezing were placed in the dark at 4°C for 12 h followed by 4 day at 22°C and the survival rates was determined. Photos were taken to record the growth phenotype before treatment and after recovery. Leaves of plants after 4°C for 24 h were collected for analysis of POD and SOD activities, MDA and soluble sugar content as described by Li et al. (2000). Seven cold responsive genes were selected for RT-PCR assay. *AtTUB2* and *AtUBQ10* of *Arabidopsis* was used as double internal reference gene. Three biological replications were carried out for each sample. The primers of the genes for RT-PCR are listed in **Supplementary Table 2**.

Statistical Analysis

Data were statistically processed by the SPSS 25.0 and Graphpad Prism 8 software. The mean value ± standard deviation (SD) of at least three replicates for each sample are presented. Statistical significance was assessed by Student's *t*-test between control and treatment.

RESULTS

Identification of ADF Gene Family Members in Wheat

A total of 25, 18, 8, 5, 12, and 11 non-redundant putative ADF genes were identified in wheat (*TaADF1–TaADF25*), *T. dicoccoides* (*TdADF1–TdADF18*), *Ae. tauschii* (*AetADF1–AetADF8*), *T. urartu* (*TuADF1–TuADF5*), *T. turgidum* (*TtADF1–TtADF12*), and barley (*HvADF1–HvADF11*), respectively (**Table 1** and **Supplementary Table 3**). Of the encoded TaADF proteins, TaADF4/8 and TaADF22 were, respectively, revealed as the shortest (132 amino acids) and longest (235 amino acids). The molecular weights (MW) of the TaADF proteins ranged from 15.30 to 25.91 kDa and the isoelectric points (pI) ranged from 4.39 to 8.74. The GRAVY values (<0) reflected the hydrophilicity of the TaADF proteins. An analysis of the instability index suggested that 16 proteins (64%) may be unstable (instability index > 40) and nine proteins (36%) are probably stable (instability index ranging from 30.91 to 38.53). The aliphatic index, which ranged from 62.52 to 79.72, indicated the thermal stability of TaADF proteins.

Analysis of the Chromosomal Locations and Duplication of *TaADF* Genes

The results of the chromosomal localization and collinear analysis revealed that the *TaADF* genes are unevenly distributed on 15 chromosomes, with the number of *TaADF* genes on each chromosome ranging from 1 (1A, 1B, 1D, 4A, 4B, 4D, 6A, 6B, and 6D) to 4 (5A and 5D) (**Figure 1A**). Nine, seven, and nine *TaADF* genes were detected in the A, B, and D sub-genomes, respectively, implying some of the *TaADF* genes in the B sub-genome may be lost during evolution. These lost ADF genes in 5B may have a redundant function with the ADF genes on 5A or 5D, with a low purifying selection, they were finally lost during evolution. We found 7 homologous gene groups with a copy on each of A, B and D sub-genomes, and 2 gene pairs had two homologous genes on A and D sub-genomes. In addition, the homologous *TaADF* genes shared high protein sequence similarity, with a range of 92.5% (*TaADF16-5A*, *TaADF18-5B*, *TaADF22-5D*) to 100% (*TaADF15-5A* and *TaADF19-5D*) (**Supplementary Table 4**). Gene duplication events affected the *TaADF* genes on 15 chromosomes. Thirty-one segmental duplications (**Figure 1B**) and one tandem duplication (**Figure 1A**) were detected. The rates of non-synonymous (Ka) and synonymous (Ks) nucleotide substitutions were calculated to evaluate the selection pressure on the *TaADF* gene duplication events (**Supplementary Table 5**). The Ka/Ks ratios for the 32 duplicated pairs were less than 1.00, implying the wheat ADF genes evolved under strong purifying selection.

Systematic Evolutionary Relationships Among ADF Members in Wheat and Eight Other Plant Species

To investigate the evolutionary relationships and characteristics of the ADF genes, 117 ADF proteins from wheat and other

TABLE 1 | Description of actin depolymerizing factor (ADF) genes identified from the wheat database.

Gene name	Gene ID	Physical Position			Information of Proteins						
		Chr	Start Position	End Position	ProteinLength	MW (Da)	pI	Instability Index	Aliphatic Index	GRAVY	Subcellular Location
TaADF1	TraesCS1A02G183100	1A	331517066	331518825	150	17310.45	5.08	42.61	72.73	−0.48	Cytoplasm
TaADF2	TraesCS1B02G197600	1B	355155044	355157167	150	17368.48	4.97	42.11	72.73	−0.50	Cytoplasm
TaADF3	TraesCS1D02G186600	1D	257474295	257475731	151	17493.70	4.97	40.05	76.75	−0.44	Cytoplasm
TaADF4	TraesCS2A02G410700	2A	668242515	668243490	132	15318.39	5.57	42.26	67.95	−0.47	Cytoplasm
TaADF5	TraesCS2A02G227300	2A	239381406	239386945	139	16142.22	5.75	38.25	62.52	−0.58	Cytoplasm
TaADF6	TraesCS2B02G255200	2B	290519507	290523695	139	16112.19	5.75	37.03	62.52	−0.58	Cytoplasm
TaADF7	TraesCS2B02G429500	2B	616934383	616935480	139	15918.01	5.87	37.09	65.25	−0.45	Cytoplasm
TaADF8	TraesCS2D02G408100	2D	523028514	523029011	132	15304.36	5.57	47.62	67.95	−0.47	Cytoplasm
TaADF9	TraesCS2D02G235200	2D	220670605	220675415	139	16168.30	5.75	42.80	65.32	−0.55	Cytoplasm
TaADF10	TraesCS4A02G071400	4A	68688963	68691716	143	16529.07	8.74	53.59	73.64	−0.35	Cytoplasm
TaADF11	TraesCS4B02G227300	4B	475810763	475813477	143	16529.07	8.74	53.59	73.64	−0.345	Cytoplasm
TaADF12	TraesCS4D02G228100	4D	387441390	387445034	143	16543.10	8.74	53.69	73.64	−0.35	Cytoplasm
TaADF13	TraesCS5A02G478800	5A	651858719	651859682	147	15800.62	4.48	35.58	79.05	−0.15	Cytoplasm
TaADF14	TraesCS5A02G416100	5A	605100341	605103635	145	16748.06	5.45	43.99	71.31	−0.42	Cytoplasm
TaADF15	TraesCS5A02G478900	5A	651982952	651984951	138	16084.13	5.65	38.53	63.62	−0.66	Cytoplasm
TaADF16	TraesCS5A02G478500	5A	651826274	651827210	200	22273.65	5.03	53.11	67.30	−0.56	Cytoplasm
TaADF17	TraesCS5B02G491800	5B	659983832	659984917	147	15824.67	4.60	34.84	79.05	−0.19	Cytoplasm
TaADF18	TraesCS5B02G491500	5B	659769899	659770854	142	15751.54	4.39	46.02	79.72	−0.32	Cytoplasm
TaADF19	TraesCS5D02G492400	5D	525658928	525661018	138	16084.13	5.65	38.53	63.62	−0.66	Cytoplasm
TaADF20	TraesCS5D02G423800	5D	483645409	483648797	145	16802.05	5.26	44.77	71.31	−0.46	Cytoplasm
TaADF21	TraesCS5D02G492300	5D	525602262	525603227	147	15886.80	4.78	30.91	75.71	−0.21	Cytoplasm
TaADF22	TraesCS5D02G491900	5D	525471415	525472578	235	25906.92	5.43	58.10	68.51	−0.50	Cytoplasm. Nucleus.
TaADF23	TraesCS6A02G247000	6A	457685808	457687435	139	15979.21	6.33	41.02	68.71	−0.43	Cytoplasm
TaADF24	TraesCS6B02G277600	6B	502483095	502484438	145	16584.01	6.33	37.25	68.55	−0.36	Cytoplasm
TaADF25	TraesCS6D02G229200	6D	320228058	320229328	139	15997.29	5.59	47.70	69.42	−0.38	Cytoplasm

Chr, Chromosome; MW, Molecular weight (Da); pI, Isoelectric point; GRAVY, Grand average of hydropathicity.

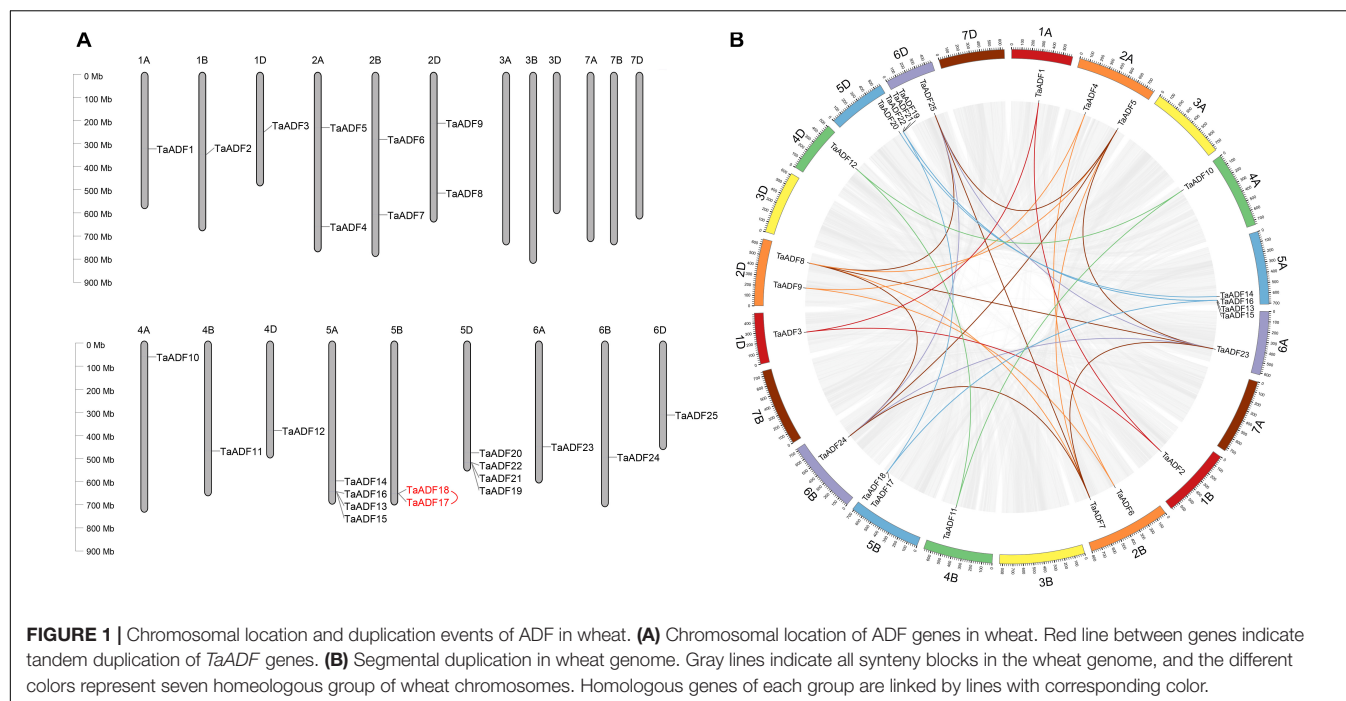


FIGURE 1 | Chromosomal location and duplication events of ADF in wheat. **(A)** Chromosomal location of ADF genes in wheat. Red line between genes indicate tandem duplication of *TaADF* genes. **(B)** Segmental duplication in wheat genome. Gray lines indicate all syntenic blocks in the wheat genome, and the different colors represent seven homeologous group of wheat chromosomes. Homologous genes of each group are linked by lines with corresponding color.

species were used to construct a phylogenetic tree (Figure 2 and Supplementary Table 3). The evolutionary relationships between wheat and eight other species were determined (Figure 2A). The phylogenetic tree revealed that the ADF genes can be classified into four main groups, with each clade consisting of 15–56 members (Figure 2B and Supplementary Table 6). More specifically, 3, 5, 6, and 11 *TaADF* are clustered in Groups I, II, III, and IV, respectively.

To more thoroughly determine the phylogenetic mechanisms of *TaADF* genes, we examined the synteny between wheat and other four gramineous species, including *Ae. tauschii*, *T. dicoccoides*, barley, and rice. A total of 16, 26, 18, and 22 orthologous gene pairs between hexaploid wheat (*T. aestivum*) and other species (*Ae. tauschii*, *T. dicoccoides*, barley, and rice) were identified (Figure 3 and Supplementary Table 7). Some collinear pairs (with eight *TaADF* genes) were identified in all of the four syntenic maps, suggesting that these orthologous pairs were relatively well conserved during the evolution of gramineous species. Each of ADF genes in 2A, 2B, 4A, 4B of tetraploid wheat showed synteny to several ones in hexaploid wheat. However, some orthologous gene pairs were only identified between chromosome 5A (or 5B) of *T. dicoccoides* and 5A of hexaploid wheat, but not 5B of hexaploid wheat, which may be due either to the quality of the genome assembly or the gene deletion or chromosomal recombination during evolution and polyploidization.

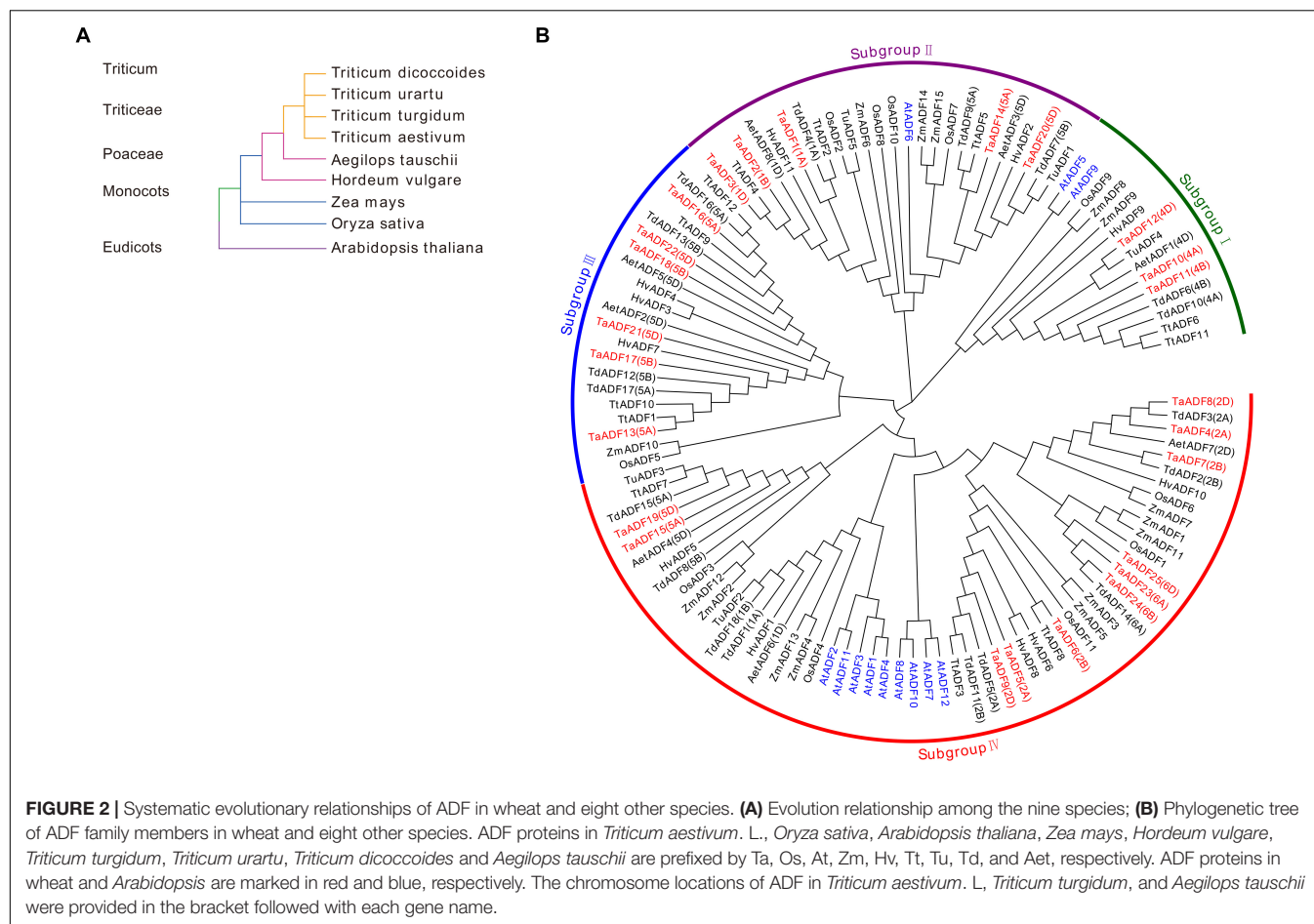
Structural Characterization of *TaADF* Genes in Wheat

The *TaADF* amino acid sequences were aligned and a phylogenetic tree was constructed using the MEGA 7.0 program (Figure 4A). A structural analysis of the *TaADF* genes indicated

that 19 *TaADF* genes have three exons, whereas the remaining genes have two exons (Figure 4B). Additionally, 19 *TaADF* genes consist of a 150-bp exon at the C-terminus and a second exon comprising 247–289 bp. Thus, the exon lengths and positions appear to be highly conserved in wheat *TaADF* genes. A total of 14 conserved motifs were identified (motifs 1–14) (Figures 4C,D). Motifs 1, 2, and 5 are three conserved regions that form the ADF-H domain in all *TaADF* proteins, whereas motif 3 is present in the N-terminus of 23 *TaADF* proteins (Figure 4C). Alignment of the predicted *TaADFs* revealed that the ADF-H domain position and actin binding sites were conserved in all of the ADFs (Supplementary Figure 1). Additionally, our analysis indicated that closely related homologous *TaADF* genes in the A, B, or D sub-genomes are usually similar regarding their structures and encoded motifs, suggesting wheat *TaADF* genes were conserved during evolution.

Analysis of *Cis*-Elements of *TaADF* Genes in Wheat

The *cis*-acting elements in the 1,500-bp upstream promoter region of the identified *TaADF* genes were predicted using the PlantCARE online program. Fifty-seven *cis*-acting elements related to cell cycle regulation, plant development, hormone responses, stress responses, and transcription are presented in Supplementary Figure 2. Many *cis*-acting elements were associated with responses to various hormones, including abscisic acid (ABRE), methyl jasmonate (CGTCA-motif and TGACG-motif), auxin (TGA-element and AuxRR-core), and gibberellin (P-box and GARE-motif). Some of the identified *cis*-acting elements may regulate the development of specific tissues such as the endosperm (GCN4-motif), seed (RY-element), and meristem (CAT-box). The promoters of 16 and 13 *TaADF*



genes included *cis*-acting elements responsive to drought (MBS) and low temperature (LTR), respectively. The presence of multiple *cis*-acting elements in the *TaADF* promoters may be indicative of the diversity in the biological functions of the encoded proteins.

Expression Profiles of *TaADF* Genes in Various Wheat Tissues

To gain insights into the *TaADF* expression patterns in diverse wheat tissues, the available RNA-seq data for various wheat tissues across different developmental stages were obtained from the Wheat eFP database (Figure 5A). The *TaADF* expression levels varied among tissues at the same growth stage. At the flag leaf stage, the TPM of *TaADF15* was 113.68, 248.09, and 302.91 in the leaf blade, root, and shoot axis, respectively. Nine genes (*TaADF4/5/6/7/8/9/23/24/25*), located on chromosome 2 or 6, were highly expressed in the anther (TPM: 239.41–823.54), but were expressed at low or undetectable levels (TPM < 1) in the leaf blade, root apical meristem, root, shoot axis, and grain. Additionally, the expression of most *TaADF* genes in specific tissues varied substantially at different growth stages. For example, *TaADF15/16/19/18/22* expression levels in the root apical meristem and root were higher at the tillering stage than at the three-leaf stage. During the four examined grain

developmental stages, the expression levels of these genes were highest and lowest at the milk grain stage and ripening stage, respectively. The tissue- and temporal-specific expression of *TaADFs* may help to clarify the complex functions of *TaADF* in various cellular processes.

The expression pattern of *TaADFs* were examined by RT-PCR in leaf and root at three-leaf stage (Supplementary Figure 3). The expression of homologous genes *TaADF15-5A* and *TaADF19-5D* were extremely high in both leaf and root, while the homologous genes *TaADF10-4A*, *TaADF11-4B*, and *TaADF12-4D*, showed significantly lower expression abundances. These results indicated that homologous *TaADF* genes in closely related clades appear to be expressed in different tissues similarity, implying they may be functionally similar.

Expression Profiles of *TaADF* Genes Under Cold Conditions

To further evaluate the potential functions of ADFs in response to low temperatures, the ADF gene expression patterns under cold acclimation and freezing conditions were analyzed based on the fragment per kilobase of transcript per million reads mapped (FPKM). Results showed that *TaADFs* had differential expression under different temperature treatment. The expression of *TaADF13/16/17/18/21/22* was induced by the cold acclimation



FIGURE 3 | Synteny analysis of *ADF* genes between wheat and other four species. Gray lines in the background and red lines between different species indicate the collinear blocks and syntenic *ADF* pairs between wheat and other species, respectively.

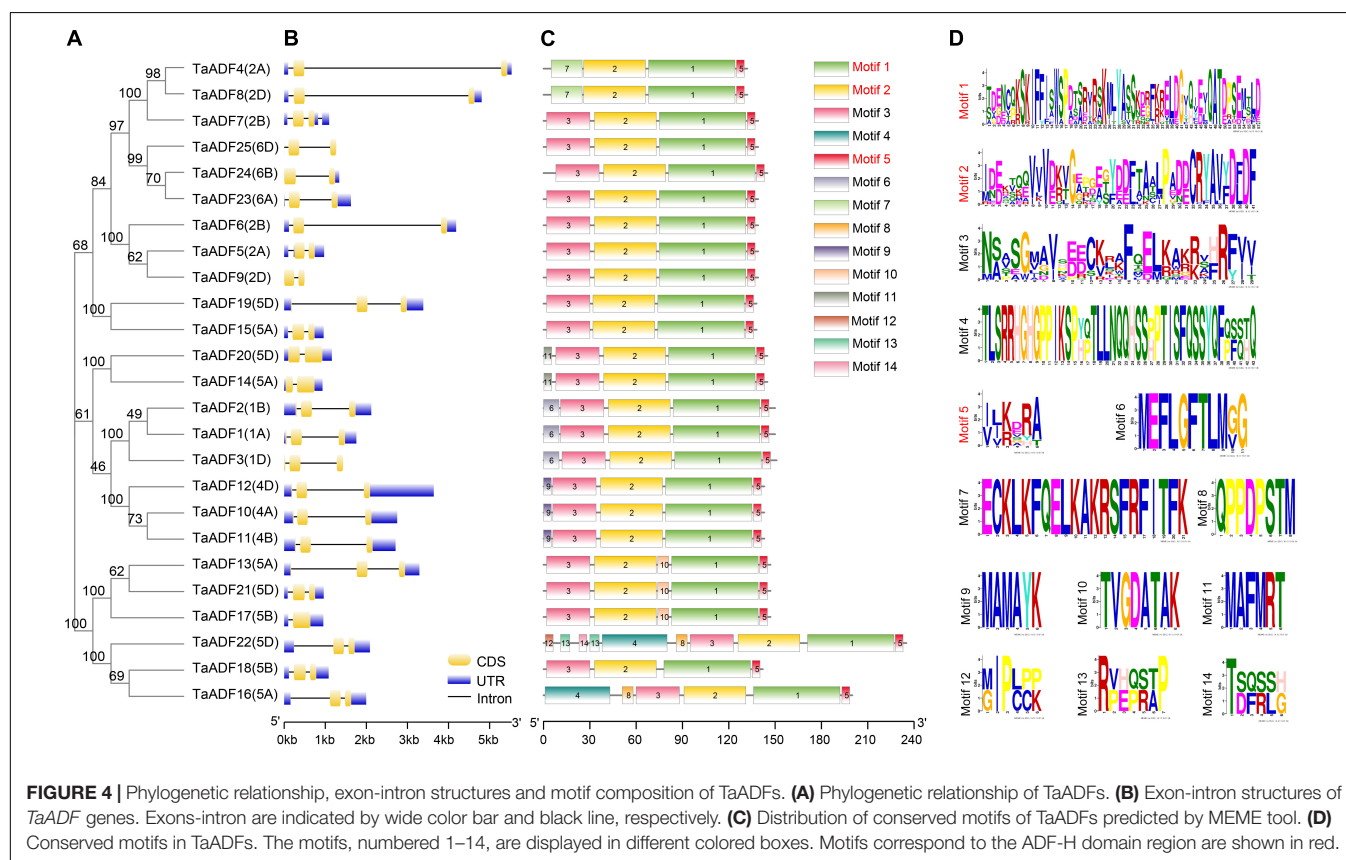
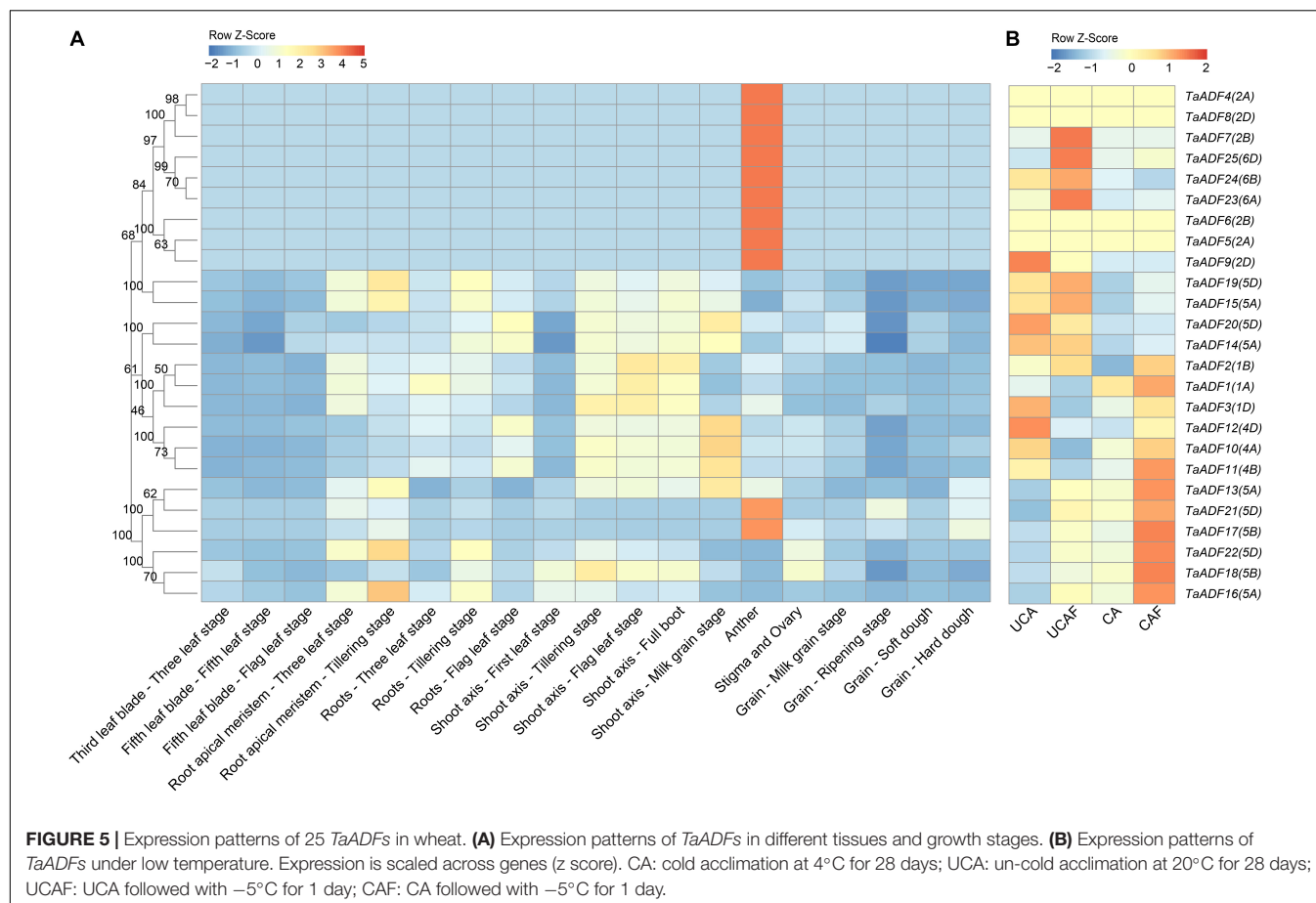


FIGURE 4 | Phylogenetic relationship, exon-intron structures and motif composition of *TaADFs*. **(A)** Phylogenetic relationship of *TaADFs*. **(B)** Exon-intron structures of *TaADF* genes. Exons-intron are indicated by wide color bar and black line, respectively. **(C)** Distribution of conserved motifs of *TaADFs* predicted by MEME tool. **(D)** Conserved motifs in *TaADFs*. The motifs, numbered 1–14, are displayed in different colored boxes. Motifs correspond to the ADF-H domain region are shown in red.

and freezing treatments, while the expression of *TaADF14/20* were decreased (**Figure 5B**). In total, seven DEGs were identified in three comparison (CA vs. UCA, CAF vs. CA and UCAF vs. UCA) (**Supplementary Table 8**). With the exception of *TaADF20*, the expression levels of all DEGs were up-regulated under the cold acclimation or freezing conditions. Significant changes to *TaADF13/20* expression were detected only in the CA vs. UCA. Up-regulated *TaADF17* expression was detected only in the CAF vs. CA. The *TaADF16/18* expression levels were up-regulated after the CA and UCA samples when exposed

to freezing stress (CAF vs. CA and UCAF vs. UCA), with a greater change in expression in the samples that did not undergo the cold acclimation process. These results suggest that *TaADF16/17/18* contribute to the freezing tolerance of wheat plants acclimated to the cold.

To verify the *TaADF* expression patterns induced by the cold acclimation and freezing conditions, the expression levels of 10 *TaADF* genes were analyzed by RT-PCR. The consistency between the RT-PCR data and the RNA-seq data was reflected by the calculated correlation coefficient ($R^2 = 0.83$)



(Supplementary Figure 4). These results confirmed the accuracy of the RNA-seq results.

Subcellular Localization of TaADFs

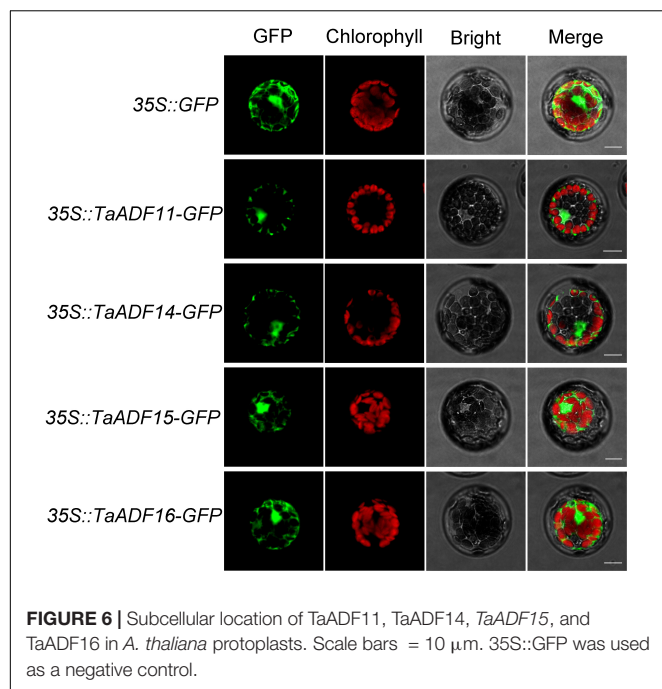
Based on the predicted subcellular localizations, all the *TaADFs* are cytoplasmic proteins, whereas *TaADF22* is present in the cytoplasm and nucleus (Table 1). To further confirm the prediction of subcellular localization of *TaADFs*, we constructed a fusion vector, which was transformed into *A. thaliana* protoplasts and observed by laser scanning confocal microscope (Figure 6). All the four proteins (*TaADF11*, *TaADF14*, *TaADF15*, and *TaADF16*) have strong fluorescence signal in transformed *A. thaliana* protoplasts, the four protein are localized in the cytoplasm and nucleus.

Overexpression of *TaADF16* Enhanced the Freezing Tolerance of *Arabidopsis*

Among the *TaADF* genes, *TaADF16* was the most highly expressed (FPKM: CA 684.5, CAF 1837.33, UCAF 932.94) and up-regulated gene (CA vs. UCA 3.72-folds, CAF vs. CA, 1.56-folds, UCAF vs. UCA, 4.12-folds) under cold acclimation or freezing treatment (Figure 5B and Supplementary Table 8). Therefore, it was functionally characterized using transgenic *Arabidopsis* plants. Three transgenic lines with

high *TaADF16* expression levels were selected for further analyses (Figures 7A,B). Before the freezing treatment, there were no obvious differences between the wild-type (WT) and *TaADF16*-overexpressing (OE) plants. Morphological changes consistent with freezing damage were detected in the WT and OE lines, but the damage was more severe in the WT plants (Figure 7C). The survival rate (%) of WT was 16.67%, whereas the survival rate of OE8, OE9, and OE11 lines were obviously higher (75.83%, 71.67, and 74.87%, respectively) after recovery (Figure 7D). The electrolyte leakage after the freezing treatment was significantly greater in the WT (70.65%) plants than in the OE lines (OE8 43.78%, OE9 42.28%, and OE11 44.87%) (Figure 7E), suggesting the cell membranes were more severely damaged in the WT plants than in the transgenic lines.

The activity of POD and SOD, as well as the soluble sugar content of the OE lines were similar to those in the WT before cold treatment, but the OE lines had higher POD and SOD activities and accumulated more soluble sugar than the WT plants after a 24-h incubation at 4°C (Supplementary Figures 5A–C). Following the cold treatment, the MDA content increased in both WT and OE lines, but MDA was significantly less abundant in the OE lines (OE8 3.38 $\mu\text{mol/g}$, OE9 2.80 $\mu\text{mol/g}$, and OE11 2.83 $\mu\text{mol/g}$) than in the WT plants (4.40 $\mu\text{mol/g}$) (Supplementary Figure 5D). Therefore, the

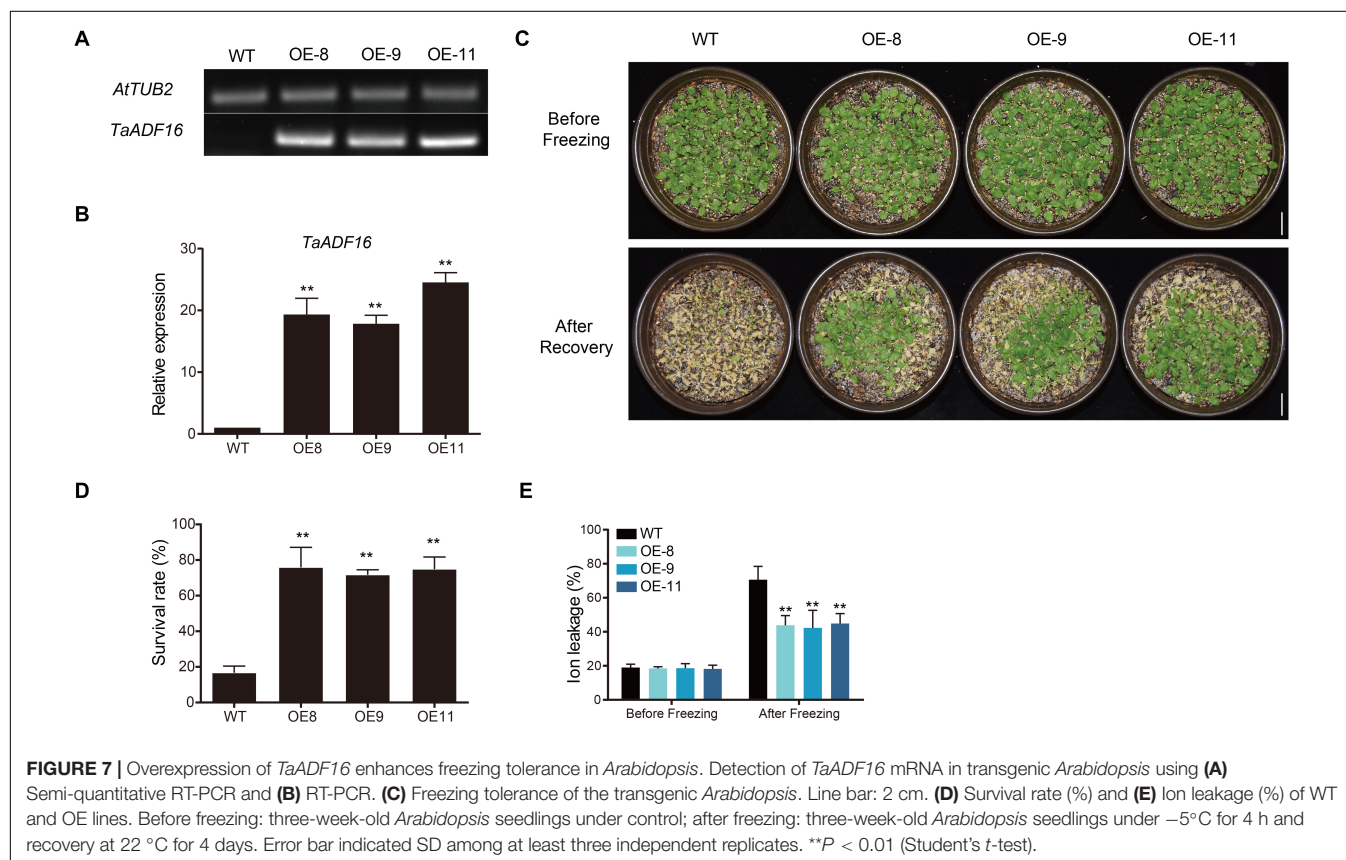


overexpression of *TaADF16* appeared to enhance the freezing tolerance of *Arabidopsis* by modulating the scavenging of ROS and by altering the osmotic regulation.

To further investigate the regulatory effects of *TaADF16* in response to low-temperature stress, seven cold-responsive genes (*CBF1*, *CBF2*, *CBF3*, *COR15A*, *COR15B*, *COR47*, and *RD22*) were selected for a RT-PCR analysis. The basal and cold-induced expression levels of these genes were higher in the OE lines than in the WT plants (**Supplementary Figure 6**). These results suggest that *TaADF16* overexpression induces the expression of cold-responsive genes, thus enhancing the cold resistance of transgenic plants.

DISCUSSION

The ADF gene family is relatively small in higher plants, with only 12, 12, 13, 11, and 14 members in rice, *Arabidopsis*, maize, tomato and poplar, respectively (Feng et al., 2006; Roy-Zokan et al., 2015; Khatun et al., 2016; Huang et al., 2020). In this study, we identified 25, 18, 12, 11, 8, and 5 ADF genes in the wheat, *T. dicoccoides*, *T. turgidum*, barley, *Ae. Tauschii*, and *T. urartu* genomes. The fact that more ADF genes were detected in wheat than in the other species may be attributed to the two rounds of polyploidization that occurred during wheat evolution (Marcussen et al., 2014). The presence of only a few ADF genes in the genomes of wheat relatives indicates that *TaADF* genes evolved after naturally occurring genomic hybridizations and fusions. The ADF gene family is believed to be structurally and functionally conserved in



plants (McCurdy et al., 2001). Analyses of the phylogenetic relationships, gene structures, and encoded motifs indicated that closely related *TaADF* homologous in the A, B, or D sub-genomes have similar exon–intron structures and encode the same conserved motifs, which is consistent with the results of an earlier study (Feng et al., 2006). Our phylogenetic tree revealed that the ADF genes in the examined monocots are clustered in four main groups, with the ADF genes in the eudicot *Arabidopsis* distributed in subgroups I, II, and V. The phylogenetic relationships among the ADF genes from selected species were consistent with those described in published reports (Khatun et al., 2016; Huang et al., 2020). Accordingly, the ADF genes in the analyzed flowering plants likely evolved from a common ancestor.

Tandem and segmental duplications are considered to be the major driving force of gene family expansions during evolution (Cannon et al., 2004). Segmental duplications were revealed in this study as the main events responsible for the evolution of the *TaADF* gene family (Figure 1B). Similar events likely occurred in maize and tomato (Khatun et al., 2016; Huang et al., 2020). Therefore, we speculate that the ADF gene families in higher plants expanded primarily because of segmental duplications. The homologous ADF genes at the branch ends of each clade of sub-genome A, B, or D are likely the putative homoalleles of genes that evolved in *Ae. tauschii*, *T. dicoccoides*, and bread wheat (Figure 2B).

Tissue- and temporal-specific expression patterns of genes in growing plants usually reflect the differences in the biological functions of gene family members as well as the cross-talk among the associated pathways (Hu et al., 2018; Zhao et al., 2019a). Nine *TaADF* genes (*TaADF4/5/6/7/8/9/23/24/25*) were highly expressed in the anther, stigma, and ovary, but were expressed at low levels in the other examined tissues (Figure 5A). Similar results were reported for *PhADF1/2* in petunia (Mun et al., 2000), *SlADF1/2/10/11* in tomato (Khatun et al., 2016), and *ZmADF1/2/7/12/13* in maize (Huang et al., 2020). The actin genes in the *Arabidopsis* genome have been divided into the vegetative class (expressed predominantly in the leaves, roots, stems, petals, and sepals) and the reproductive class (highly expressed in pollen) (Meagher et al., 1999). Because ADF proteins interact with actin, researchers also classified the ADF genes into vegetative and pollen-specific groups (Mun et al., 2000). We predict that the *TaADF4/5/6/7/8/9/23/24/25* genes clustered in subgroup IV belong to the reproductive class, whereas the other *TaADF* genes are grouped in the vegetative class. The *TaADF14/15/19/20* genes were more highly expressed than the other *TaADF* genes in the vegetative class, implying these four *TaADF* genes are important for wheat growth.

In this study, we detected dynamic changes to *TaADF* expression levels during various plant growth and developmental stages. In tomato, *SlADF1/5/7/9* are highly expressed in immature fruit, whereas *SlADF3/11* expression levels peak in the mature fruit stage (Khatun et al., 2016). In the current study, in developing wheat grains, the expression of most *TaADF* genes peaked in the milk grain stage, markedly decreased in the ripening stage, and then increased in the

soft dough and hard dough stages. These findings suggest the encoded *TaADF* proteins have similar regulatory effects on actin filaments in wheat plants. However, there were no obvious expression-level trends common to all *TaADF* genes during the development of other tissues (e.g., leaf, root apical meristem, root, and shoot axis). These observations have compelled us to investigate the complex ADF regulatory mechanisms underlying wheat growth.

Previous studies confirmed that the expression of ADF gene is induced in plants exposed to low temperatures (Kerr and Carter, 1990; Danyluk et al., 1996; Ouellet et al., 2001; Tang et al., 2016). The reorganization regulated by ADFs may influence various cytoskeletal-associated cell processes. In response to low-temperature stress, microtubules are more easily depolymerized in cold-acclimated rye root tip cells than in non-acclimated cells, and this depolymerization enhances the freezing tolerance of the root tips (Kerr and Carter, 1990). We identified six *TaADF* genes (*TaADF13/16/17/18/21/22*) with significantly up-regulated expression under cold acclimation or freezing conditions, which is consistent with the cold-induced changes in *SlADF2/11* expression in tomato (Khatun et al., 2016). However, *TaADF20* expression was down-regulated by cold stress. This phenomenon might be explained by the antagonistic relationships among ADFs. For example, in *Arabidopsis*, *AtADF9* adversely affects *AtADF1* by regulating its ability to depolymerize actin, whereas the opposite effect is observed when *AtADF9* and *AtADF1* are ectopically expressed in tobacco cells (Tholl et al., 2011). Notably, *TaADF16/18/22* expression levels were low during all examined wheat growth stages, but were highly up-regulated by low-temperature stress. Accordingly, to cope with cold conditions, *TaADF* gene expression is induced in wheat plants, thereby increasing the remodeling of actin. These findings reveal the complexity of the *TaADF* regulatory mechanism under cold conditions. Osmotic and ROS homeostasis is essential for plant cold tolerance (Zuo et al., 2019). In our study, the overexpression of *TaADF16* increased the freezing tolerance of *Arabidopsis* plants, likely because of the positive effects on ROS scavenging and osmotic regulation (Supplementary Figure 5). Furthermore, the expression of cold-responsive genes was induced in the transgenic *Arabidopsis* (Supplementary Figure 6), suggesting that *TaADF16* may regulate cold tolerance by interacting with ICE-CBF-related genes. However, to more comprehensively characterize the relationship between the remodeling of the actin cytoskeleton and wheat responses to cold stress, additional ADF genes may need to be identified and functionally annotated.

CONCLUSION

In this study, we identified 25 ADF genes in wheat. Based on the protein sequence alignment, 117 ADFs from wheat and the other analyzed species were clustered into four main groups. Segmental duplications during evolution were important for the expansion of the *TaADF* gene family. Analyses of the phylogenetic relationships, gene structures, and encoded motifs suggested that *TaADF* were conserved during evolution.

The tissue- and temporal-specific expression patterns of *TaADF* genes were revealed in this study. Nine genes preferentially expressed in the anther (*TaADF4/5/6/7/8/9/23/24/25*) are likely associated with pollen development. Additionally, we identified seven differentially expressed *TaADF* genes after low-temperature treatments. Specifically, the expression of homologous genes *TaADF16/18/22* were considerably induced by cold stress, implying these genes are critical for the freezing tolerance of wheat. Overexpression of *TaADF16* substantially increased the tolerance of transgenic plants to freezing stress because of the associated effects on the cell membrane and ROS homeostasis, as well as the CBF/DREB pathway genes. These results provide new insights into the regulatory functions of TaADF proteins related to wheat responses to low temperature, and provides candidate gene resources for breeding new wheat varieties with enhanced freezing tolerance.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: NCBI GEO database: GSE135474 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135474>).

AUTHOR CONTRIBUTIONS

YZ and XY conceived the project, set the scientific objectives, and revised the manuscript. KX, SiZ, HL, WW, and ShZ conducted the experiments and analyzed the data. KX wrote the manuscript. All authors discussed the results as well as read and approved the final manuscript for publication.

FUNDING

This work was funded by the National Key Research and Development Program of China (2017YFD0101000), The Technology Innovation of Winter Wheat of Science and Technology Planning Project of Hebei Province (16226320D), and The Innovation Fund Project for Graduate Student of Hebei Province (CXZZBS2019101).

REFERENCES

- Allard, A., Bouzid, M., Betz, T., Simon, C., Abou-Ghali, M., Lemiere, J., et al. (2020). Actin modulates shape and mechanics of tubular membranes. *Sci. Adv.* 6:eaz3050. doi: 10.1126/sciadv.aaz3050
- Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., Castro, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 40, 597–603. doi: 10.1093/nar/gks400
- Bamburg, J. R., Harris, H. E., and Weeds, A. G. (1980). Partial purification and characterization of an actin depolymerizing factor from brain. *FEBS Lett.* 121, 178–182. doi: 10.1016/0014-5793(80)81292-0
- Burgos-Rivera, B., Ruzicka, D. R., Deal, R. B., McKinney, E. C., King-Reid, L., and Meagher, R. B. (2008). ACTIN DEPOLYMERIZING FACTOR9 controls

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.618984/full#supplementary-material>

Supplementary Figure 1 | Multiple alignments of TaADFs. The solid arrow and open arrow indicate the binding site of G-actin (ADFs shown in red) and F-actin (all the 25 ADFs), respectively. The red solid line indicates the ADF-H domain position.

Supplementary Figure 2 | Analysis of *cis*-acting elements in TaADFs.

Supplementary Figure 3 | Analysis of expression profiles of TaADF genes in leaf (A) and root (B) by RT-PCR. Error bar indicated SD among at least three independent replicates, and the values differed significantly when $P < 0.05$. Varied letters meant a significant difference.

Supplementary Figure 4 | Confirmation of RNA-seq results by RT-PCR. (A) Expression patterns of *TaADFs* by RNA-seq and RT-PCR. (B) Correlation between RNA-seq and RT-PCR for selected TaADF genes. The relative expression of *TaADFs* in crown of three leaf stage seedlings (TL) was set as the control. Three biological replicates for each sample were performed and bars represent the SD. CA: cold acclimation at 4°C for 28 days; UCA: un-cold acclimation at 20°C for 28 days; UCAF: UCA followed with −5°C for 1d; CAF: CA followed with −5°C for 1d.

Supplementary Figure 5 | Analysis of SOD activity (A), POD activity (B), MDA content (C), and soluble sugar content (D) in WT and OE lines of *Arabidopsis*. Before cold: three-week-old *Arabidopsis* seedlings under control (22°C); after cold: three-week-old *Arabidopsis* seedlings under 4°C for 24 h. Error bar indicates SD among at least three independent replicates. ** $P < 0.01$ (Student's *t*-test).

Supplementary Figure 6 | Expression levels of seven cold response genes in WT and OE lines before or after cold stress. Before cold: three-week-old *Arabidopsis* seedlings under control; after cold: three-week-old *Arabidopsis* seedlings under 4°C for 24 h. Error bar indicates SD among at least three independent replicates. ** $P < 0.01$ (Student's *t*-test).

Supplementary Table 1 | The data of genome sequences used in this study.

Supplementary Table 2 | Specific primers used in this study.

Supplementary Table 3 | The information of ADF protein of different plant species in this study.

Supplementary Table 4 | Sequence similarities among the TaADF proteins.

Supplementary Table 5 | Ka/Ks ratios of collinearity *TaADF* genes in wheat.

Supplementary Table 6 | The number of ADF proteins of different subgroups in phylogenetic tree.

Supplementary Table 7 | ADF collinear gene pairs between wheat and other selected species.

Supplementary Table 8 | The log2 (fold change) of differently expressed genes (DEGs) under cold stresses.

development and gene expression in *Arabidopsis*. *Plant Mol. Biol.* 68, 619–632. doi: 10.1007/s11103-008-9398-1

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* 4:10. doi: 10.1186/1471-2229-4-10

Chen, C. Y., Wong, E. I., Vidali, L., Estavillo, A., Hepler, P. K., Wu, H. M., et al. (2002). The regulation of actin organization by actin-depolymerizing factor in elongating pollen tubes. *Plant Cell* 14, 2175–2190. doi: 10.1105/tpc.003038

Chou, K. C., and Shen, H. B. (2010). Plant-mPLOC: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS One* 5:e11335. doi: 10.1371/journal.pone.0011335

- Daher, F. B., and Geitmann, A. (2012). Actin depolymerizing factors ADF7 and ADF10 play distinct roles during pollen development and pollen tube growth. *Plant Signal. Behav.* 7, 879–881. doi: 10.4161/psb.20436
- Danyluk, J., Carpentier, E., and Sarhan, F. (1996). Identification and characterization of a low temperature regulated gene encoding an actin-binding protein from wheat. *FEBS Lett.* 389, 324–327. doi: 10.1016/0014-5793(96)00599-6
- Dong, C. H., Xia, G. X., Hong, Y., Ramachandran, S., Kost, B., and Chua, N. H. (2001). ADF proteins are involved in the control of flowering and regulate F-actin organization, cell expansion, and organ growth in *Arabidopsis*. *Plant Cell* 13, 1333–1346. doi: 10.1105/tpc.13.6.1333
- Drobak, B. K., Franklin-Tong, V. E., and Staiger, C. J. (2004). The role of the actin cytoskeleton in plant cell signaling. *New Phytol.* 163, 13–30. doi: 10.1111/j.1469-8137.2004.01076.x
- Du, J., Wang, X., Dong, C. H., Yang, J. M., and Yao, X. J. (2016). Computational study of the binding mechanism of actin-depolymerizing factor 1 with actin in *Arabidopsis thaliana*. *PLoS One* 11:e0159053. doi: 10.1371/journal.pone.0159053
- Feng, Y., Liu, Q., and Xue, Q. (2006). Comparative study of rice and *Arabidopsis* actin-depolymerizing factors gene families. *J. Plant Physiol.* 163, 69–79. doi: 10.1016/j.jplph.2005.01.015
- Fowler, J. E., and Quatrano, R. S. (1997). Plant cell morphogenesis: plasma membrane interactions with the cytoskeleton and cell wall. *Annu. Rev. Cell Dev. Biol.* 13, 697–743. doi: 10.1146/annurev.cellbio.13.1.697
- Fu, Y., Duan, X., Tang, C., Li, X., Voegele, R. T., Wang, X., et al. (2014). TaADF7, an actin-depolymerizing factor, contributes to wheat resistance against *Puccinia striiformis* f. sp. *tritici*. *Plant J.* 78, 16–30. doi: 10.1111/tjp.12457
- Gungabissoon, R. A., Jiang, C. J., Drobak, B. K., Maciver, S. K., and Hussey, P. J. (1998). Interaction of maize actin-depolymerizing factor with actin and phosphoinositides and its inhibition of plant phospholipase C. *Plant J.* 16, 689–696. doi: 10.1046/j.1365-313x.1998.00339.x
- Henty, J. L., Bledsoe, S. W., Khurana, P., Meagher, R. B., Day, B., Blanchoin, L., et al. (2011). *Arabidopsis* actin depolymerizing factor 4 modulates the stochastic dynamic behavior of actin filaments in the cortical array of epidermal cells. *Plant Cell* 23, 3711–3726. doi: 10.1105/tpc.111.090670
- Hu, C. H., Wei, X. Y., Yuan, B., Yao, L. B., Ma, T. T., Zhang, P. P., et al. (2018). Genome-wide identification and functional analysis of NADPH oxidase family genes in wheat during development and environmental stress responses. *Front. Plant Sci.* 9:906. doi: 10.3389/fpls.2018.00906
- Huang, J., Sun, W., Ren, J., Yang, R., Fan, J., Li, Y., et al. (2020). Genome-wide identification and characterization of actin-depolymerizing factor (ADF) family genes and expression analysis of responses to various stresses in *Zea mays* L. *Int. J. Mol. Sci.* 21:1751. doi: 10.3390/ijms21051751
- Huang, Y. C., Huang, W. L., Hong, C. Y., Lur, H. S., and Chang, M. C. (2012). Comprehensive analysis of differentially expressed rice actin depolymerizing factor gene family and heterologous overexpression of OsADF3 confers *Arabidopsis thaliana* drought tolerance. *Rice* 5:33. doi: 10.1186/1939-8433-5-33
- International Wheat Genome Sequencing Consortium (IWGSC) (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345:1251788. doi: 10.1126/science.1251788
- Kerr, G. P., and Carter, J. V. (1990). Relationship between freezing tolerance of root-tip cells and cold stability of microtubules in Rye (*Secale cereale* L. cv Puma). *Plant Physiol.* 93, 77–82. doi: 10.1104/pp.93.1.77
- Khatun, K., Robin, A. H., Park, J. I., Kim, C. K., Lim, K. B., Kim, M. B., et al. (2016). Genome-wide identification, characterization and expression profiling of ADF family genes in *Solanum lycopersicum* L. *Genes* 7:79. doi: 10.3390/genes7100079
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Li, H. S., Sun, Q., Zhao, S. J., Zhang, W. H., Chen, C. L., and Hong, Y. Z. (2000). *The Experiment Principle and Technique on Plant Physiology and Biochemistry*. Beijing: Higher Education Press.
- Li, X. B., Xu, D., Wang, X. L., Huang, G. Q., Luo, J., Li, D. D., et al. (2010). Three cotton genes preferentially expressed in flower tissues encode actin-depolymerizing factors which are involved in F-actin dynamics in cells. *J. Exp. Bot.* 61, 41–53. doi: 10.1093/jxb/erp280
- Liu, D., Sun, J., Zhu, D., Lyu, G., Zhang, C., Liu, J., et al. (2019). Genome-wide identification and expression profiles of late embryogenesis-abundant (LEA) genes during grain maturation in wheat (*Triticum aestivum* L.). *Genes* 10:696. doi: 10.3390/genes10090696
- Lopez, I., Anthony, R. G., Maciver, S. K., Jiang, C. J., Khan, S., Weeds, A. G., et al. (1996). Pollen specific expression of maize genes encoding actin depolymerizing factor-like proteins. *Proc. Natl. Acad. Sci. U.S.A.* 93, 7415–7420. doi: 10.1073/pnas.93.14.7415
- Maciver, S. K., and Hussey, P. J. (2002). The ADF/cofilin family: actin-remodeling proteins. *Genome Biol.* 3, reviews3007.1–reviews3007.12. doi: 10.1186/gb-2002-3-5-reviews3007
- Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K. S., et al. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345, 1250092. doi: 10.1126/science.1250092
- McCurdy, D. W., Kovar, D. R., and Staiger, C. J. (2001). Actin and actin-binding proteins in higher plants. *Protoplasma* 215, 89–104. doi: 10.1007/Bf01280306
- Meagher, R. B., McKinney, E. C., and Vitale, A. V. (1999). The evolution of new structures: clues from plant cytoskeletal genes. *Trends Genet.* 15, 278–284. doi: 10.1016/s0168-9525(99)01759-x
- Mun, J. H., Yu, H. J., Lee, H. S., Kwon, Y. M., Lee, J. S., Lee, I., et al. (2000). Two closely related cDNAs encoding actin-depolymerizing factors of petunia are mainly expressed in vegetative tissues. *Gene* 257, 167–176. doi: 10.1016/S0378-1119(00)00412-1
- Ouellet, F., Carpentier, E., Cope, M. J., Monroy, A. F., and Sarhan, F. (2001). Regulation of a wheat actin-depolymerizing factor during cold acclimation. *Plant Physiol.* 125, 360–368. doi: 10.1104/pp.125.1.360
- Porter, K., and Day, B. (2016). From filaments to function: the role of the plant actin cytoskeleton in pathogen perception, signaling and immunity. *J. Integr. Plant Biol.* 58, 299–311. doi: 10.1111/jipb.12445
- Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., et al. (2018). The transcriptional landscape of polyploid wheat. *Science* 361:eaar6089. doi: 10.1126/science.aar6089
- Roy-Zokan, E. M., Dyer, K. A., and Meagher, R. B. (2015). Phylogenetic patterns of codon evolution in the ACTIN-DEPOLYMERIZING FACTOR/COFILIN (ADF/CFL) gene family. *PLoS One* 10:e0145917. doi: 10.1371/journal.pone.0145917
- Sengupta, S., Mangu, V., Sanchez, L., Bedre, R., Joshi, R., Rajasekaran, K., et al. (2019). An actin-depolymerizing factor from the halophyte smooth cordgrass, *Spartina alterniflora* (SaADF2), is superior to its rice homolog (OsADF2) in conferring drought and salt tolerance when constitutively overexpressed in rice. *Plant Biotechnol. J.* 17, 188–205. doi: 10.1111/pbi.12957
- Smertenko, A. P., Jiang, C. J., Simmons, N. J., Weeds, A. G., Davies, D. R., and Hussey, P. J. (1998). Ser6 in the maize actin-depolymerizing factor, ZmADF3, is phosphorylated by a calcium-stimulated protein kinase and is essential for the control of functional activity. *Plant J.* 14, 187–193. doi: 10.1046/j.1365-313x.1998.00107.x
- Song, G., Zhang, R., Zhang, S., Li, Y., Gao, J., Han, X., et al. (2017). Response of microRNAs to cold treatment in the young spikes of common wheat. *BMC Genomics* 18:212. doi: 10.1186/s12864-017-3556-2
- Staiger, C. J., Gibbon, B. C., Kovar, D. R., and Zonia, L. E. (1997). Profilin and actin-depolymerizing factor: modulators of actin organization in plants. *Trends Plant Sci.* 2, 275–281. doi: 10.1016/S1360-1385(97)86350-9
- Sung, S., and Amasino, R. M. (2005). Remembering winter: toward a molecular understanding of vernalization. *Annu. Rev. Plant Biol.* 56, 491–508. doi: 10.1146/annurev.arplant.56.032604.144307
- Tang, C. L., Deng, L., Chang, D., Chen, S. T., Wang, X. J., and Kang, Z. S. (2016). TaADF3, an actin-depolymerizing factor, negatively modulates wheat resistance against *Puccinia striiformis*. *Front. Plant Sci.* 6:1214. doi: 10.3389/fpls.2015.01214
- Tholl, S., Moreau, F., Hoffmann, C., Arumugam, K., Dieterle, M., Moes, D., et al. (2011). *Arabidopsis* actin-depolymerizing factors (ADFs) 1 and 9 display antagonist activities. *FEBS Lett.* 585, 1821–1827. doi: 10.1016/j.febslet.2011.05.019
- Thomashow, M. F. (2001). So what's new in the field of plant cold acclimation? Lots! *Plant Physiol.* 125, 89–93. doi: 10.1104/pp.125.1.89
- Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3

- Wang, H. Y., Wang, J., Gao, P., Jiao, G. L., Zhao, P. M., Li, Y., et al. (2009). Down-regulation of *GhADF1* gene expression affects cotton fibre properties. *Plant Biotechnol. J.* 7, 13–23. doi: 10.1111/j.1467-7652.2008.00367.x
- Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Zhang, B., Hua, Y., Wang, J., Huo, Y., Shimono, M., Day, B., et al. (2017). TaADF4, an actin-depolymerizing factor from wheat, is required for resistance to the stripe rust pathogen *Puccinia striiformis* f. sp. *tritici*. *Plant J.* 89, 1210–1224. doi: 10.1111/tpj.13459
- Zhao, Y., Ma, R. Y., Xu, D. L., Bi, H. H., Xia, Z. L., and Peng, H. R. (2019a). Genome-wide identification and analysis of the AP2 transcription factor gene family in wheat (*Triticum aestivum* L.). *Front. Plant Sci.* 10:1286. doi: 10.3389/fpls.2019.01286
- Zhao, Y., Zhou, M., Xu, K., Li, J., Li, S., Zhang, S., et al. (2019b). Integrated transcriptomics and metabolomics analyses provide insights into cold stress response in wheat. *Crop J.* 7, 857–866. doi: 10.1016/j.cj.2019.09.002
- Zheng, Y., Xie, Y., Jiang, Y., Qu, X., and Huang, S. (2013). *Arabidopsis* actin-depolymerizing factor7 severs actin filaments and regulates actin cable turnover to promote normal pollen tube growth. *Plant Cell* 25, 3405–3423. doi: 10.1105/tpc.113.117820
- Zhou, M., Zheng, S., Liu, R., Lu, J., Lu, L., Zhang, C., et al. (2019). Genome-wide identification, phylogenetic and expression analysis of the heat shock transcription factor family in bread wheat (*Triticum aestivum* L.). *BMC Genomics* 20:505. doi: 10.1186/s12864-019-5876-x
- Zuo, Z. F., Kang, H. G., Park, M. Y., Jeong, H., Sun, H. J., Song, P. S., et al. (2019). Zoysia japonica MYC type transcription factor ZjICE1 regulates cold tolerance in transgenic *Arabidopsis*. *Plant Sci.* 289:110254. doi: 10.1016/j.plantsci.2019.110254

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Xu, Zhao, Zhao, Liu, Wang, Zhang and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The *WtmsDW* Locus on Wheat Chromosome 2B Controls Major Natural Variation for Floret Sterility Responses to Heat Stress at Booting Stage

Million F. Erena¹, Iman Lohraseb¹, Isabel Munoz-Santa^{1,2}, Julian D. Taylor¹, Livinus C. Emebiri^{3,4} and Nicholas C. Collins^{1*}

¹ School of Agriculture Food and Wine, The University of Adelaide, Adelaide, SA, Australia, ² Department of Statistics and Operations Research, University of Valencia, Valencia, Spain, ³ Graham Centre for Agricultural Innovation, Charles Sturt University, Wagga Wagga, NSW, Australia, ⁴ New South Wales Department of Primary Industries, Wagga Wagga, NSW, Australia

OPEN ACCESS

Edited by:

Peter Stephen Baenziger,
University of Nebraska System,
United States

Reviewed by:

Sivakumar Sukumaran,
International Maize and Wheat
Improvement Center, Mexico
Mary Guttieri,
Agricultural Research Service,
United States Department
of Agriculture, United States

*Correspondence:

Nicholas C. Collins
nicholas.collins@adelaide.edu.au

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 14 December 2020

Accepted: 17 February 2021

Published: 29 March 2021

Citation:

Erena MF, Lohraseb I, Munoz-Santa I, Taylor JD, Emebiri LC and Collins NC (2021) The *WtmsDW* Locus on Wheat Chromosome 2B Controls Major Natural Variation for Floret Sterility Responses to Heat Stress at Booting Stage. *Front. Plant Sci.* 12:635397. doi: 10.3389/fpls.2021.635397

Heat stress at booting stage causes significant losses to floret fertility (grain set) and hence yield in wheat (*Triticum aestivum* L.); however, there is a lack of well-characterized sources of tolerance to this type of stress. Here, we describe the genetic analysis of booting stage heat tolerance in a cross between the Australian cultivars Drysdale (intolerant) and Waagan (tolerant), leading to the definition of a major-effect tolerance locus on the short arm of chromosome 2B, *Wheat thermosensitive male sterile Drysdale/Waagan (WtmsDW)*. *WtmsDW* offsets between 44 and 65% of the losses in grain set due to heat, suggesting that it offers significant value for marker-assisted tolerance breeding. In lines lacking the *WtmsDW* tolerance allele, peaks in sensitivity were defined with reference to auricle distance, for various floret positions along the spike. Other (relatively minor) floret fertility response effects, including at the *Rht-D1* dwarfing locus, were considered likely escape artifacts, due to their association with height and flowering time effects that might interfere with correct staging of stems for heat treatment. Heat stress increased grain set at distal floret positions in spikelets located at the top of the spike and increased the size of spikelets at the base of the spike, but these effects were offset by greater reductions in grain set at other floret positions. Potentially orthologous loci on chromosomes 1A and 1B were identified for heat response of flowering time. The potential significance of these findings for tolerance breeding and further tolerance screening is discussed.

Keywords: wheat, heat tolerance, male sterility, floret sterility, auricle distance, QTL

INTRODUCTION

Heat stress reduces yields of wheat in most global production environments, and the situation is worsening with climate change (Asseng et al., 2015).

Elevated temperatures accelerate development, senescence, and water use, reducing the opportunity to accumulate biomass and therefore yield (Asseng et al., 2011; Hunt et al., 2018). Heat waves (> 30°C for one to several days) during sensitive reproductive development stages

also impact directly on grain set (floret fertility) and grain development. Wheat has two periods of sensitivity to the floret sterility effects: during booting (flag leaf sheath extending, 1–2 weeks before anthesis), and 2–3 days before anthesis (Saini and Aspinall, 1982; Tashiro and Wardlaw, 1990a; Craufurd et al., 2013; Prasad and Djanaguiraman, 2014; Barber et al., 2017). Heat stress during early grain filling reduces the weight of individual grains (Tashiro and Wardlaw, 1990b; Stone and Nicolas, 1995). Grain filling heat stress can also affect grain physical and biochemical traits determining processing characteristics and end-use quality (Stone and Nicolas, 1998; Wrigley, 2007).

Modeling has indicated that every further degree Celsius rise in mean global temperature would result in a 6% loss in wheat yields worldwide (Asseng et al., 2015). In Australia, it was estimated that heat shocks during reproductive development reduced grain number and individual grain weight by 3.6 and 18.1%, respectively, translating to a yield loss of 20.8%, in the mid-maturing wheat cultivar Janz, over the period 1985–2017 (Ababaei and Chenu, 2020).

To help limit these yield losses, tolerant wheat cultivars able to withstand heat waves during reproductive development could be grown. However, breeder's efforts to identify heat tolerance in the field are hampered by the unpredictable timing and intensity of natural heat events and the narrow developmental windows of sensitivity. Selection using molecular markers would be more convenient than phenotypic selection. However, this strategy would firstly require identification of loci controlling major variation for heat tolerance.

Various efforts have been made to identify grain filling heat tolerance quantitative trait loci (QTL) in hexaploid wheat, involving transfers of potted plants into a growth chamber set at high temperatures, at 7–10 days after anthesis (Mohammadi et al., 2008; Mason et al., 2010, 2011; Shirdelmoghanloo et al., 2016; Guo et al., 2020; Lu et al., 2020). However, relatively few studies have targeted floret fertility responses to heat stress applied at booting or at around anthesis (Barber et al., 2017). Barber et al. (2017) identified one weak tolerance QTL for booting stage, plus two tolerance QTL for anthesis stage that were associated with a locus for dwarfing (*Rht-D1*) and flowering time (*Ppd-D1*), respectively.

Accordingly, in the current study, we performed QTL analysis on floret fertility responses to heat stress applied at booting, in a cross between the Australian wheat cultivars Drysdale and Waagan. An initial study indicated Drysdale was relatively sensitive to heat stress at this stage and Waagan tolerant (Erena, 2018). We have also used this population to identify grain filling heat tolerance QTL (Shirdelmoghanloo et al., 2016), providing the opportunity to compare tolerance QTL for the two developmental stages.

MATERIALS AND METHODS

Plant Genetic Materials and Markers

An initial experiment was undertaken to discover QTL for heat responsiveness of traits in a Drysdale × Waagan F₁-derived doubled haploid (DH) population, by applying a brief heat stress

at booting stage (“DH QTL experiment”). The population of 144 lines, and the genetic map of 551 nonredundant marker loci, has previously been described (Shirdelmoghanloo et al., 2016).

After identifying a strong floret fertility heat tolerance locus on chromosome 2B (*WtmsDW*), KASP™ assays were utilized to map this region in more detail and develop *WtmsDW* near-isogenic lines. KASP™ assays were run using an automated SNPLine system and Kraken™ software (DNA LGC Limited, London, United Kingdom). In addition to using a pre-existing KASP assay for the *Ppd-B1* gene, three other KASP markers in the region were developed, including one based on the SNP *wsnp_JD_c3732_4781170* from the wheat 9k Illumina iSelect SNP array (Cavanagh et al., 2013) and *AHW_DW_001* and *AHW_DW_014* based on SNPs between Drysdale and Chinese Spring identified using the DAWN genomics tool (Watson-Haigh et al., 2018; **Supplementary Table 1**). KASP markers were scored on the DH lines to confirm their locations (**Supplementary Figure 1**).

A Drysdale × Waagan recombinant inbred line (RIL) population was made for developing *WtmsDW* near-isogenic lines. The RIL population was derived by single-seed descent from F₂ plants of a Drysdale × Waagan cross. The four KASP markers were used to identify an F₆ RIL plant that was heterozygous for the *WtmsDW* region. This was then used to derive a *WtmsDW*-heterozygous F₈ plant by two further rounds of single-seed descent with marker selection. The progeny of this single plant were screened to identify three plants homozygous for each allele type. These were then allowed to self-pollinate to establish seed stocks of the six near-isogenic lines (NILs): NIL-T-1, NIL-T-2, and NIL-T-3 (Waagan allele, tolerant) and NIL-I-1, NIL-I-2, and NIL-I-3 (Drysdale allele, intolerant). Four progeny of the same *WtmsDW*-heterozygous F₈ plant (sibs of the selected NILs), together with the Drysdale and Waagan parents, had been subjected to genomic profiling using DArTSeq™. These data were used to identify the chromosome segments segregating in this material and hence differing between the NILs. Genomic locations of the DArTSeq markers in the wheat IWGSC v1.0, Chinese Spring reference genome sequence had been determined using BLAST searches.

Greenhouse Conditions, Heat Treatment, and Data Collection

To phenotype the Drysdale × Waagan DH population in the DH QTL experiment, plants were grown in the Plant Accelerator facility at the Waite Campus of the University of Adelaide, using procedures similar to Shirdelmoghanloo et al. (2016). Plants were sown on 16th March 2014 and grown in a naturally lit evaporatively cooled greenhouse compartment, where max-day/min-night temperatures averaged 20/17°C and day/night relative humidity averaged 68/76% throughout growth. The greenhouse temperature reached 27.2°C once during the treatment period due to high outside temperatures. The experiment was arranged in three sections of the greenhouse. Each section comprised a rectangular array of pots, each of them indexed by its row and column. The experiment was designed in a split plot layout with four blocks (replicates). Genotypes

(DH lines and parents) were randomly allocated to main plots comprising pairs of adjacent pots in rows and treatments (control and heat) to subplots comprising the two pots within main plots. Each plant was heat treated when the main stem reached a certain growth stage, defined by the distance between the auricles (collars) of the flag leaf and the next leaf down (auricle distance, AD; Jagadish et al., 2014). Treatments began on the day the AD on the main stem was closest to 3 cm (for replicates 1 and 3) or 9 cm (for replicates 2 and 4), i.e., during mid-booting (growth stages Z41–Z46; Zadoks et al., 1974). AD was also measured to the nearest 0.5 cm on the day of heat treatment, on both the main stem and the most advanced tiller, and these two stems were marked with different color tags. Plants were moved to a walk-in growth chamber (Conviron BDW120) set at 14 h day-length and 37/27°C day/night temperature. Maximum temperature was held for 8 h, with 3-h ramping periods either side (Supplementary Figure 2A). Day/night relative humidity was around 60/80% in the chamber. While in the chamber, pots sat in trays containing ~2 cm of water to minimize drought stress. Plant movement in and out of the chambers each day was done at the start of the night cycle. After 3 days, plants were moved back to the greenhouse to complete their development. Traits (Table 1) were

TABLE 1 | Traits measured in the DH QTL experiment.

Trait abbreviation	Description
Day.AD	No. days from sowing to targeted auricle distance (pre-heat)
Day.Anth	No. days from sowing to first anthesis (extrusion of first anther) on the main stem
Day.ADtoAnth	Trait “Day.Anth” minus trait “Day.AD”
AwnEm.PreH	Length of emerged awns at targeted auricle distance (pre-heat; cm)
AD.Mat	Auricle distance at maturity (cm)
Ht.Mat	Plant height at maturity, soil level to bottom of spike (cm)
SpkL.Mat	Spike length at maturity, to glumes of terminal spikelet (cm)
AwnL.Mat	Length of awns above terminal spikelet at maturity (cm)
UndvSplt.SpK	No. of basal under-developed spikelets per spike at maturity. Under-developed spikelets were defined as those with awn length < 50% that of the spikelets from the middle of the spike.
NoSplt.SpK	Total spikelet no. per spike at maturity
GrNoSplt.1&2.Top	Grains per developed spikelet, floret positions 1 and 2, top third of spike
GrNoSplt.1&2.Mid	Grains per developed spikelet, floret positions 1 and 2, middle third of spike
GrNoSplt.1&2.Bot	Grains per developed spikelet, floret positions 1 and 2, bottom third of spike
GrNoSplt.>2.Top	Grains per developed spikelet, floret positions > 2, top third of spike
GrNoSplt.>2.Mid	Grains per developed spikelet, floret positions > 2, middle third of spike
GrNoSplt.>2.Bot	Grains per developed spikelet, floret positions > 2, bottom third of spike
GrNoSplt.SpK	Grain number per developed spikelet, across all floret and spike positions

Day.AD, Day.Anth, and Day.ADtoAnth were only scored in main stems.

scored on the tagged stems. The AD of the tagged advanced tiller averaged 1.6 and 6.5 cm when the main stem AD was 3 and 9 cm, respectively. Thus, four developmental stages of the measured stems were defined for analysis of heat responses: 1.6, 3, 6.5 and 9 cm AD.

Transfer of the Drysdale × Waagan DH lines to the growth chamber involved a change in day length (10.3 h in the greenhouse vs. 14 h in the growth chamber) as well as temperature. To clarify which environmental factor was relevant to *WtmsDW*, a second experiment was performed using the *WtmsDW* NILs (“*WtmsDW* validation experiment”). Plants were sown on 22nd July 2020 in the greenhouse, with a natural day length of 11.5 h at booting. Reach-in chambers (Conviron PGC20) containing a mixture of halogen incandescent lamps and fluorescent tubes were used to apply four treatments in which heat stress and day length were varied (Supplementary Figures 2A–D). As only two chambers were available, the earliest developing plants were used for the two treatments involving heat, and the slightly later plants were then used for the two treatments involving no heat. For the former, the main stem was tagged and used for data collection, while in the latter, either a main stem or advanced tiller was used. Each plant was treated when AD on the tagged stem was 6 cm. Treatment duration was 2 or 3 days, for treatments with and without heat stress, respectively. After treatment, plants were moved back to the greenhouse to complete their development. Plant arrangement in the greenhouse was a completely randomized design, with 19 to 41 plants (replicates) ultimately being treated per allele/treatment combination.

Phenotypic Modeling for DH QTL Experiment

Each trait in each stem was analyzed using the following linear mixed model:

$$y = X\tau + Z_g g + Z_b u_b + Z_m u_m + e \quad (1)$$

where y is the vector of observations, τ is the vector of fixed effects containing the terms to capture the treatment by genotype effects (DH or check lines) with associated design matrix X , g is the vector of random genetic effects with design matrix Z_g , u_b is the vector of random block effects with design matrix Z_b , u_m is the vector of random main plot effects with design matrix Z_m , and e is the residual error.

The joint distribution of (g, u_b, u_m, e) was assumed to be Gaussian with zero mean and variance covariance matrix:

$$\begin{bmatrix} G_g(\gamma_g) & 0 & 0 & 0 \\ 0 & \sigma_b^2 I_b & 0 & 0 \\ 0 & 0 & \sigma_m^2 I_m & 0 \\ 0 & 0 & 0 & R(\phi) \end{bmatrix}$$

where γ_g , σ_b^2 , σ_m^2 , and ϕ are unknown variance parameters associated with the genetic effects, block variance, main plot variance, and residual error, respectively. The I matrix denotes the identity matrix. Pots in the greenhouse were divided in three sections so $R(\phi)$ was assumed to be the direct sum of

$\bigoplus_{s=1}^3 \sigma_s^2 \Sigma(\rho_{cs}) \otimes \Sigma(\rho_{rs})$. The parameter σ_s^2 denotes the error variance and $\Sigma(\rho_{rs}) \otimes \Sigma(\rho_{cs})$ refers to an autoregressive process of order one in the column and row directions in section s . This is a plausible model to account for the correlation between errors due to the neighboring pots within each section. Most importantly, $\mathbf{G}_g(\mathbf{y}_g)$ represents the variance covariance matrix of the genetic effects for the DH lines only. There were three treatments so $\mathbf{g}$ was partitioned into $\mathbf{g}_c$, $\mathbf{g}_3$, and $\mathbf{g}_9$ for the control, heat applied at the AD closest to 3 and 9 cm genetic effects, respectively. The variance matrix $\mathbf{G}_g(\mathbf{y}_g)$ was then assumed to be:

$$\begin{bmatrix} \sigma_c^2 & \sigma_{c3} & \sigma_{c9} \\ \sigma_{c3} & \sigma_3^2 & \sigma_{39} \\ \sigma_{c9} & \sigma_{39} & \sigma_9^2 \end{bmatrix} \otimes \mathbf{I}_g$$

For all traits measured, tolerance genetic effects at the targeted development stages were derived based on the conditional distribution of the heat genetic effects given the control genetic effects (Lemerle et al., 2006). For instance, the tolerance genetic effects obtained from the application of the heat treatment at the AD closest to 3 cm, $\mathbf{g}_{t3}$, is distributed as follows:

$$\mathbf{g}_{t3} = \mathbf{g}_3 | \mathbf{g}_c \sim N(E(\mathbf{g}_3 - \beta_3 \mathbf{g}_c) = 0, \sigma_{t3}^2 \mathbf{I}_g) \quad (2)$$

where $\beta_3 = \rho_{c3} \frac{\sigma_3}{\sigma_c}$, $\rho_{c3} = \sigma_{c3} / \sqrt{(\sigma_c^2 \sigma_3^2)}$ and $\sigma_{t3}^2 = \sigma_3^2 (1 - \rho_{c3}^2)$.

Following (2), $\mathbf{g}_{t3}$ can be viewed as residuals from a random regression of $\mathbf{g}_3$ against $\mathbf{g}_c$ with intercept zero and slope β_3 . Genotypes with large positive residuals have higher tolerance to heat stress than an average genotype while large negative residuals suggest poor tolerance. A similar derivation can be used to obtain tolerance effects for the application of the heat treatment at the AD closest to 9 cm, i.e., $\mathbf{g}_{t9}$.

Benefits of defining tolerance effects in this way are that the effects are reported in the original unit of measurement and are independent of the control genetic effects. This avoids the problem inherent in some other commonly used response indexes, such as the Heat Susceptibility Index (Mason et al., 2010) that tend to be influenced by control *per se* performance. Tolerance effects defined in this way were also used by McDonald et al. (2015) and Mahjourimajd et al. (2016).

All models (1) were diagnostically assessed to ensure that the assumptions of normality and homoscedasticity of errors were satisfied and, where appropriate, traits were transformed. The methods described in Gilmour et al. (1997) were used to account for possible spatial trends across the experimental layout in the glasshouse. Where appropriate, the linear mixed model (1) was adapted to include linear row and column terms in the fixed part of the model or random row or column effects. The significance of the correlations of the error section terms was also assessed.

The method of residual maximum likelihood (REML) was used for variance parameter estimation (Patterson and Thompson, 1971). The best linear unbiased predictions (eBLUPs) for each line at each treatment were extracted from the model, and the tolerance effects were derived as residuals from the random regressions using the REML estimates. All analyses were conducted in the R environment (R Core Team, 2020) using the ASReml-R software (Butler et al., 2017).

QTL Mapping

For each of the traits, the eBLUPs for the DH lines under control and the two different heat conditions, as well as the eBLUPs for heat tolerance calculated from (2), were used to conduct QTL analysis. QTL mapping was performed using the approach of Shirdelmoghanloo et al. (2016). Following simple interval mapping, candidate QTL were used as co-factors for composite interval mapping (CIM), setting the minimum co-factor proximity to 30 cM and maximum step size to 10 cM. Putative QTL were considered significant if they exceeded a genome-wide LOD threshold of 1.8 calculated using the adjusted Bonferroni-corrected p value with significance level $\alpha=0.05$ (Li and Ji, 2005). To assist interpretation, QTL effects linked within ~ 30 cM were grouped to the same numbered QTL locus. All QTL analyses were conducted using the statistical computing environment GenStat version 16 (Payne, 2009; VSN International, 2020). Markers in four genomic locations showed segregation distortion (on linkage groups 2B1, 3B1, 5A2, and 6B2; χ^2 test, $p < 0.01$), but these were not located at any of the QTL reported in the current study.

Relationship of Treatment Stage to Heat-Induced Sterility and Its Interaction With Stem Type, Floret Position, and Genotype at WtmsDW and Rht Loci

The markers most closely associated with *WtmsDW* effects (*Ppd-B1*, *wsnp_Ex_c5412_9565527*, *wsnp_JD_c3732_4781170*, *wsnp_RFL_Contig4483_5312236*) were used to infer the *WtmsDW* allele (Drysdale or Waagan) carried by each DH line (marker recombinants being excluded). *Rht-B1* and *Rht-D1* alleles carried by DH lines were known from scores of diagnostic KASP markers for these genes (Shirdelmoghanloo et al., 2016). Fertility was plotted against AD length on the day of treatment and 3rd-order polynomial trend curves fitted using Microsoft Excel.

Relationships to Other Fertility Loci

Other fertility loci responsive to temperature and other environmental factors were considered for their potential relationships to *WtmsDW*. These included loci described in wheat (Kuchel et al., 2007; Mason et al., 2010, 2011, 2013; Pinto et al., 2010; Zheng et al., 2010; Pandey et al., 2014; Sharma et al., 2016; Shirdelmoghanloo et al., 2016; Barber et al., 2017; Bhusal et al., 2017; Pradhan et al., 2019; Guo et al., 2020; Lu et al., 2020; Selva et al., 2020 and references therein), durum (*Triticum durum* Desf.) (El Hassouni et al., 2019), barley (*Hordeum vulgare* L.) (Romagosa et al., 1999; Malosetti et al., 2004), and rice (*Oryza sativa* L.) (Yu et al., 2017; Zhu et al., 2017; Fan and Zhang, 2018 and references therein; Khilaimongkhon et al., 2019; Cao et al., 2020; Nubankoh et al., 2020). Marker sequences were accessed from Cavanagh et al. (2013), GrainGenes¹, Gramene², NCBI³, and

¹<https://wheat.pw.usda.gov/GG3/>

²<http://gramene.org/>

³<https://www.ncbi.nlm.nih.gov/>

Diversity Arrays Technology⁴. Marker sequences were located in the wheat IWGSC v1.0, Chinese Spring reference genome sequence using an in house BLAST tool, or in the rice Nipponbare IRGSP Reference sequence 1.0 by BLAST search at the Rice Genome Program site⁵. To establish further wheat-rice genomic interval relationships, gene sequences were accessed through the Rice Genome Program site and the DAWN tool, and homologues located in the respective genomes by BLAST search.

RESULTS

Trait Responses

In the DH QTL experiment, all traits measured after heat treatment responded to heat in the DH lines, except for number of spikelets per spike (**Figure 1**). Heat decreased the number of spikelets at the bottom of the spike that were classified as underdeveloped (UndvSplt.SpK), by up to 28%. Heat consistently decreased grain set in the lowest two floret positions in the spikelets (GrNoSplt.1&2), but it increased grain set in the upper floret positions (GrNoSplt.>2) for stems treated at the earlier developmental stages (1.6 and 3 cm AD). However, the overall effect of heat at these early stages on grain set (GrNoSplt.SpK; **Figure 1A**) was still negative because the > 2 positions produced far fewer seeds than the lower two floret positions (e.g., 0.15 vs. 0.94 grains per spikelet, in the top third of the spike, under heat). Heat accelerated the time from sowing to first anthesis (Day.Anth), by up to 5% (3.1 days). It increased final spike length (SpkL.Mat) in stems exposed to heat at the earlier stages, by up to 7.3%, but decreased it by 0.6% for the stems exposed at the latest stage. This was due to responses in rachis internode length because total spikelet number per spike was unaffected. Heat decreased auricle distance and height at maturity (AD.Mat and Ht.Mat), with the effects being the greatest for the stems exposed to heat at later stages. Heat decreased awn length at maturity (AwnL.Mat), with the effects being the greatest for the stems exposed to heat at the earlier stages (**Figure 1B**).

Quantitative Trait Loci

Overall, 221 QTL effects were defined for *per se* traits (in control or heat) and 54 QTL effects defined for heat tolerance effects (responses) (**Supplementary Table 1**). The QTL effects were grouped into 33 genomic locations (QTLx designations). Of these, 16 coincided with loci previously identified in this population (Shirdelmoghanloo et al., 2016), while 17 were new (QTL30 to QTL45). As previously reported, this population segregated for major height effects at the *Rht-D1* and *Rht-B1* loci but was relatively uniform for flowering time (largest effect 2.9 days at QTL18 on chromosome 4B).

Six loci were defined for heat responses of floret fertility (**Table 2**). QTL36 on the short arm of chromosome 2B had the strongest tolerance effect, controlling up to 43% of the variation, with tolerance deriving from the tolerant parent Waagan. Its

per se fertility effects were almost exclusively observed under heat. It was also the most consistently expressed fertility response locus, showing an effect for 18 out of the 24 tested floret position and treatment-stage combinations (**Supplementary Table 2**). QTL36 also showed an effect for awn length *per se*. QTL36 effects were found to arise from sterility in the male reproductive organs (data not shown). Accordingly, we named this major QTL36 heat tolerance locus *wheat thermosensitive male sterile Drysdale/Waagan* (WtmsDW). Additional greenhouse experiments to characterize WtmsDW effects on pollen and anther development will be described in a separate paper.

The other five loci for floret fertility response (QTL18, 19, 32, 39 and 43) were weaker and less consistently expressed than WtmsDW (**Table 2** and **Supplementary Table 2**). They were associated with various developmental traits (**Supplementary Table 2**), suggesting these fertility effects may have been developmental artifacts rather than due to genuine tolerance (see section “Discussion”).

For other (non floret fertility) traits, heat-response effects were observed at 13 loci (**Table 3**).

Interactions of WtmsDW Heat Tolerance Expression With Treatment Stage, Stem Type, Floret Position, and Rht Genotype

Sterility in main stems and tillers of DH mapping lines showed similar response curves, to heat applied at the various stem developmental stages, either in the lines carrying the Drysdale or Waagan alleles at WtmsDW (**Supplementary Figure 3**). Therefore, data from main stem and tillers were combined for subsequent analysis.

Doubled haploid lines carrying different allele combinations at the *Rht-B1* and *Rht-D1* height loci were also compared (**Supplementary Figure 4**). The two semi-dwarf types and the tall showed similar response patterns, while intolerance in the double-dwarf types peaked at a shorter AD. Therefore, the double-dwarfs were excluded from the analysis to compare floret positions (**Figure 2** and **Supplementary Figure 5**).

In the lowest two floret positions in the spikelets (positions 1 and 2), intolerance peaked at AD ~5.5 cm in the middle of the spike, while at the top and bottom of the spike, intolerance peaked later, at around 6 and 8 cm, respectively (**Figure 2A**). Floret positions > 2 in the spikelets peaked in intolerance at approximately 11, 11, and 12.5 cm AD, in the top, middle, and bottom of the spike, respectively (**Figure 2B**).

The DH lines carrying the Waagan WtmsDW allele maintained high levels of fertility across all stages where heat stress was applied (**Supplementary Figures 5C,D**). Thus, it was unlikely that WtmsDW exerted its fertility effects by merely altering the AD vs. spike developmental stage relationship, resulting in escape.

Further Mapping in the WtmsDW Region and Relationships to Other Fertility Loci

Based on DH lines that were nonrecombinant for markers spanning the WtmsDW region (33.7–93.9 cM), heat tolerance genetic effects proved to be a good predictor of WtmsDW

⁴<https://www.diversityarrays.com/>

⁵http://rice.plantbiology.msu.edu/analyses_search_blast.shtml

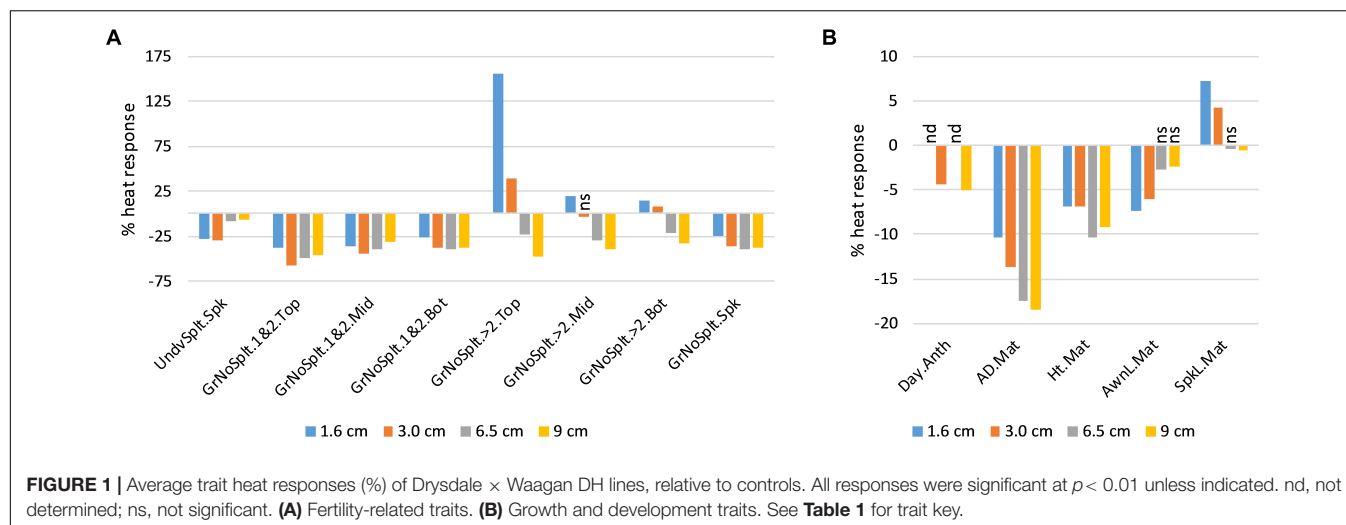


TABLE 2 | Loci for heat responses of floret fertility identified in the DH QTL experiment.

Locus	Linkage group	Position (cM)	Peak markers	Tolerance allele ^a	LOD	R^2	Additive effect
QTL32	1B	66.1	<i>wsnp_Ex_c58292_59652859</i>	D	3.8	8.6	0.01
QTL36 ^b (<i>WtmsDW</i>)	2B1	74.3–84.8	<i>Ppd-B1</i> ; <i>wsnp_JD_c3732_4781170</i> ; <i>wsnp_RFL_Contig4483_5312236</i>	W	18	43	0.37
QTL39 ^b	3B2	80.9–105.9	<i>wsnp_Ex_c9594_15882022</i> ; <i>wsnp_Ex_rep_c101457_86818160</i>	W	5.0	15	0.11
QTL18	4B	127.5	<i>wsnp_Ku_c11570_18860306</i>	D	4.3	9.3	0.01
QTL19 ^b (<i>Rht-D1</i>)	4D	0.0–2.9	<i>Rht-D1</i> ; <i>wsnp_CAP11_c356_280910</i>	D or W	6.8	12	0.18
QTL43 ^b	7A2	43.2–67.0	<i>wsnp_Ex_c2268_4251636</i> ; <i>wsnp_Ex_c12102_19361467</i>	W	4.0	6.6	0.14

^aD, Drysdale; W, Waagan.

^bFloret fertility tolerance effects were observed at these loci for multiple floret type/treatment timing combinations; figures in the last three columns are for the strongest effect.

allele type. For 68 out of the 81 nonrecombinants, positive and negative values correctly indicated the presence of the Waagan and Drysdale alleles, respectively. Inconsistencies were not clearly associated with any particular *Rht-B1/Rht-D1* genotype combination (**Supplementary Figure 1**). Accordingly, *WtmsDW* genotype was inferred for the remaining (recombinant) DH lines, which then allowed *WtmsDW* to be mapped as a single point locus. It was located to a 5.3-cM marker interval (**Figure 3**), corresponding to the physical interval 76.82 to 95.76 Mb on chromosome 2B in the IWGSC v1.0, Chinese Spring reference genome sequence (**Supplementary Figure 1**).

The DH line WW28450 carried a spontaneous deletion on chromosome 2B from 104.24 Mb upwards (**Supplementary Figure 1**), essentially covering the whole of the short arm, including the *WtmsDW* locus (**Figure 3**). This line was phenotypically intolerant to heat stress but fully fertile under control conditions (**Supplementary Figure 1**).

Previously described loci on wheat chromosome 2B influencing fertility responses to environmental factors were considered for their potential relationships to *WtmsDW*. Mason et al. (2010) reported a QTL for response (Heat Susceptibility Index (HSI)) of kernel number per main spike to post-anthesis heat stress, between markers *gwm148* and *barc200*, which is in

the vicinity of *WtmsDW* (**Figure 3**). Zheng et al. (2010) described interactions of kernel number per square meter with several environmental covariates including cumulative degree-days during the 6 days around meiosis, associated with the markers *gwm429* and *gwm374*, which is also close to *WtmsDW* (**Figure 3**). Barber et al. (2017) reported a relatively weak QTL affecting interaction of heat stress during early booting with floret fertility, associated with the marker *gwm120*. However, *gwm120* is ~40 cM from *WtmsDW* on the long arm of chromosome 2B (**Figure 3**). Sharma et al. (2016) described a HSI effect of grain number per spike, associated with markers *1161184* and *1097543*, calculated by comparing late vs. timely sown field trials. However, these markers are > 40 cM from *WtmsDW* on the long arm (**Figure 3**).

The *Wtms1*, *TmsBS20T*, and *Wptms2* male sterility loci on chromosome 2B have been defined in the context of hybrid breeding research. The *Wtms1* and *TmsBS20T* loci are expressed if it is colder than 10°C during spike development. These loci have been mapped 4.8 and 4.5 cM from marker *gwm374*, respectively (Xing et al., 2003; Ru et al., 2015), which places them at least ~10 cM from *WtmsDW* (**Figure 3**). The *Wptms2* locus expressed sterility in late sowings (Guo et al., 2006a,b), thus requiring long days and/or high temperature for expression. It

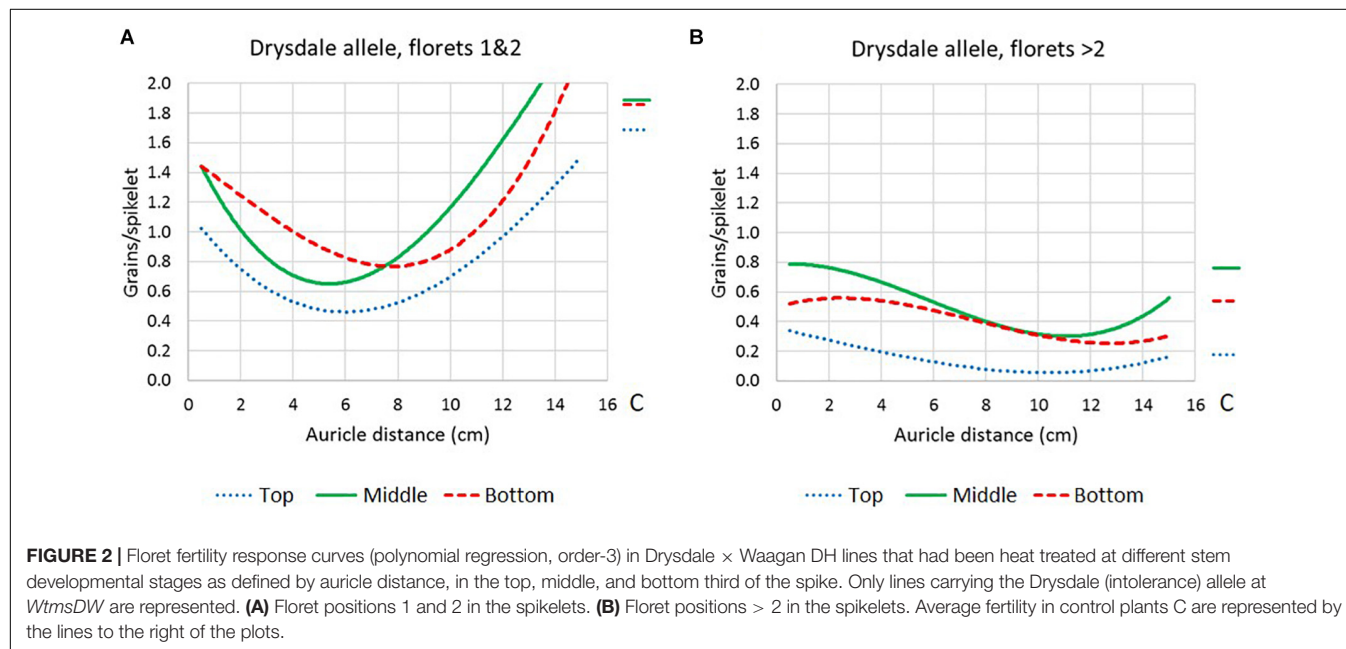
TABLE 3 | Heat-response QTL effects for traits other than floret fertility identified in the DH QTL experiment.

Locus	Trait	Linkage group	Position (cM)	Tolerance allele ^a	Treatment stage (AD) ^b	LOD	R ²	Additive effect
QTL30	Day.Anth	1A1	97.0	W	3	3.5	10.0	0.16
QTL33	Day.Anth	1B	135.2	D	9	3.5	8.9	0.74
QTL29	Day.ADtoAnth	7B	56.9	D	3	6.9	18.8	0.79
QTL15	AD.Mat	4A2	41.6	D	1.6 and 9	4.3	1.5	0.01
QTL15	Ht.Mat	4A2	41.6	D	1.6	4.1	2.1	0.01
QTL17 (<i>Rht-B1</i>)	AD.Mat	4B	83.9	W	9	47.0	44.5	0.01
QTL17 (<i>Rht-B1</i>)	Ht.Mat	4B	83.9	W	1.6	37.5	41.2	0.01
QTL18	AD.Mat	4B	141.3	W	9	4.9	1.8	0.01
QTL19 (<i>Rht-D1</i>)	AD.Mat	4D	0.0	D	9	46.0	41.9	0.01
QTL19 (<i>Rht-D1</i>)	Ht.Mat	4D	0.0	D	1.6	39.3	44.8	0.01
QTL28	AD.Mat	7A2	33.7	W	9	5.8	3.0	0.01
QTL34	Ht.Mat	2A	147.0–160.2	D	3 and 6.5	4.3	12.1	1.56
QTL29	AwnL.Mat	7B	45.8	D	6.5	3.7	9.4	0.05
QTL41	AwnL.Mat	5D2	40.2	W	3	3.5	11.3	0.01
QTL9	SpkL.Mat	2D4	7.6	D	1.6	3.6	9.2	0.01
QTL17 (<i>Rht-B1</i>)	UndvSplt.SpK	4B	83.9	D	3 and 9	16.9	26.3	0.01
QTL18	UndvSplt.SpK	4B	135.5	D	3 and 9	8.7	11.6	0.01
QTL19 (<i>Rht-D1</i>)	UndvSplt.SpK	4D	0.0	W	3 and 9	10.5	13.4	0.01
QTL25	UndvSplt.SpK	6A	74.7	D	3 and 9	4.1	4.5	0.07
QTL5.2	UndvSplt.SpK	2A	93.6	D	1.6 and 6.5	3.7	9.6	0.09

Effects are grouped by trait type: anthesis-related, height-related, awn or spike length-related, and underdeveloped lower spikelets. See **Table 1** for trait key.

^aW, Waagan; D, Drysdale.

^bTreatments giving QTL effects, based on auricle distance (AD, cm) at time of heat treatment. Where QTL effects were detected at two treatment times, figures in the last three columns are for the strongest effect.

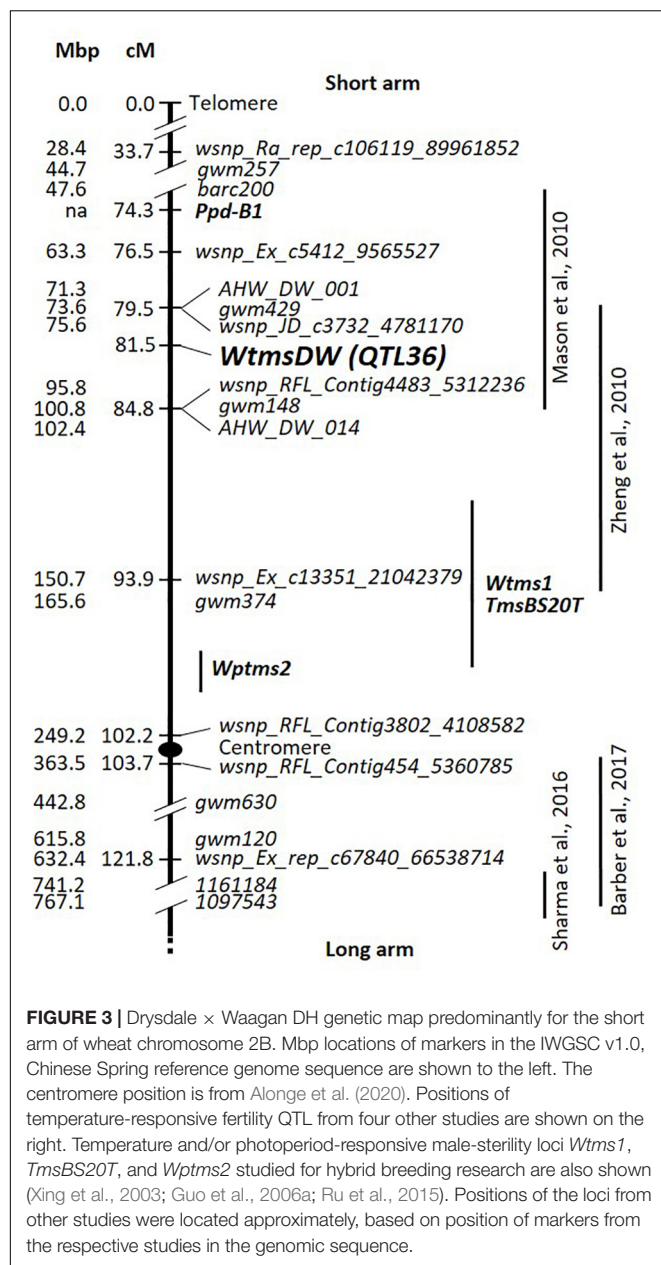


was mapped 6.9 cM below *gwm374* (Guo et al., 2006a), placing it ~20 cM from *WtmsDW* (**Figure 3**).

In rice, BLAST searches with gene sequences established that there were two genomic regions related to the *WtmsDW* interval of wheat: on chromosome 3 (11.13–12.71 Mb; Nipponbare IRGSP Reference sequence 1.0) and chromosome 7 (12.33–22.39 Mb). The *qHTB3-2* QTL influencing floret fertility

responses to heat stress at booting (Zhu et al., 2017) overlapped with the chromosome 3 interval. However, the reported rice QTL interval was relatively large (~10 Mb; 12.33–22.39 Mb) and only overlapped for ~400 kb of the ~1.58 Mb, corresponding to the *WtmsDW* interval.

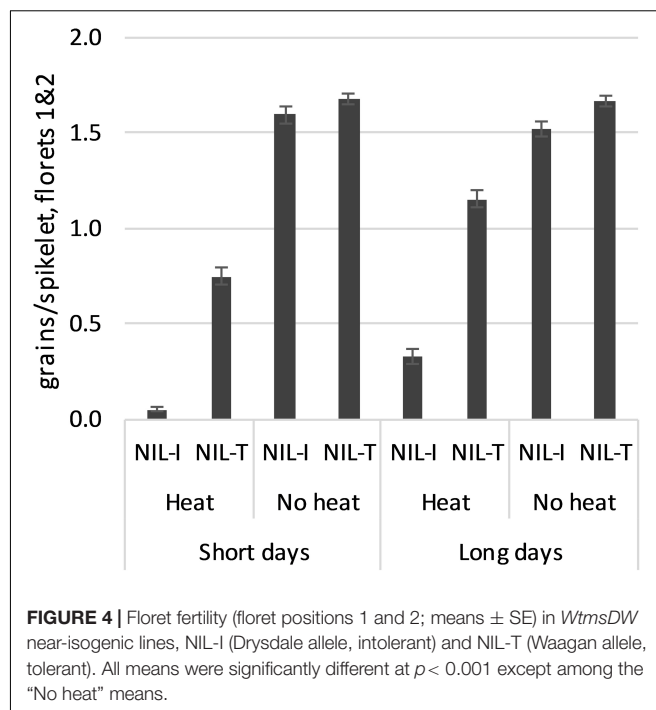
A grain yield locus responsive to temperature during heading in barley (Romagosa et al., 1999; Malosetti et al., 2004) was



located between markers *Rbc2* and *ABG002* on chromosome 2H. This corresponded to 44.36–55.88 Mb on wheat chromosome 2B, which is distal of *WtmsDW*.

WtmsDW Validation Experiment

The three NILs selected for each *WtmsDW* allele type behaved similarly (not shown), so were regarded as one for the purposes of data presentation: “NIL-I” with the intolerance allele from Drysdale, and “NIL-T” with the tolerance allele from Waagan, respectively. In the absence of heat stress, fertility was unaffected by day length (9 h vs. 14 h) in either of the NILs (**Figure 4**). Heat treatment reduced fertility in both NILs, but more so in NIL-I than in NIL-T. Heat stress reduced fertility more in both NILs under short days than under long days, although the relative



difference between the NILs remained similar (**Figure 4**). In other words, while fertility may have responded to day-length, this response was independent of *WtmsDW*. These results validated *WtmsDW* as a floret fertility heat tolerance locus and showed that its effects were not day-length dependent.

DArTSeqTM genomic profiling indicated that the NIL lines were 96% identical, differing only for a segment of chromosome 2B carrying *WtmsDW* (64.23–247.52 Mb) and small segments on chromosomes 1A and 3D which did not carry any of the other reported floret fertility heat-response QTL listed in **Table 2**. These data further supported the assertion that floret fertility effects observed in the validation experiment were due to *WtmsDW*.

DISCUSSION

WtmsDW Significance and Potential Applications

WtmsDW is a locus controlling major natural variation for male sterility in response to high temperatures at booting stage in wheat. Variation between the two Australian wheat cultivars Drysdale and Waagan was used to define *WtmsDW*. Drysdale (Hartog*3/Quarrion) was released in 2002, and Waagan (Janz/24IBWSN-244) was released in 2007, both in NSW. The *WtmsDW* alleles present in Drysdale (intolerant) and Waagan (tolerant) were inherited from Hartog and Janz, respectively (marker data not shown). Hartog and Janz have been particularly popular cultivars and (in addition to Drysdale) have been used extensively as parents in Australian wheat breeding. This suggests good potential for *WtmsDW* closest flanking KASP markers *wsnp_JD_c3732_4781170* and *AHW_DW_014* (**Supplementary Table 1**) to be used in current Australian

breeding programs, either to select for heat tolerance or against intolerance. In the validation experiment using *WtmsDW* NIL lines, the tolerance allele offset 44–65% of the losses in grain set due to heat stress otherwise experienced in lines carrying the intolerance allele (**Figure 4**), suggesting that the use of these KASP markers could lead to substantial yield benefits. However, *WtmsDW* tolerance would need to be evaluated in a number of genetic backgrounds and multiple field environments over a number of years to understand its true value for breeding.

Heat tolerance QTL were previously described in the *WtmsDW* region of chromosome 2B (**Figure 3**). Mason et al. (2010) identified a QTL in this region for heat response (Heat Susceptibility Index) of grain number per spike, in a cross between the Australian spring cultivar Halberd and the winter wheat Cutter, for heat treatments commencing 10 days after anthesis. However, in our experience, heat treatments applied at this stage in a range of genetic material has not affected grain number (Shirdelmoghanloo et al., 2016, and unpublished data). Zheng et al. (2010) identified a QTL for grain number per square meter that interacted with several environmental variables, including cumulative degree-days over 25°C in the 6 days around meiosis, across 12 field environments, in the winter wheat cross Arche × Réclat. Meiosis occurs in stems with an AD around 0–5 cm (Browne et al., 2018), which is around when *WtmsDW* effects were expressed (**Supplementary Figure 5**). Although further work will be needed to confirm whether these loci are equivalent to *WtmsDW*, these other studies hint that *WtmsDW* selection might be applicable across a wide range of breeding programs.

The Drysdale × Waagan DH population has also been screened for responses to a 3-day heat stress applied at 10 days after anthesis (Shirdelmoghanloo et al., 2016). Loci on chromosomes 3B and 6B influenced the ability to maintain both higher grain weight and flag leaf chlorophyll content under heat, suggesting that grain filling heat tolerance may have a mechanism relating to lower rates of heat-enhanced senescence in both the leaves and grains. By contrast, *WtmsDW* was located on chromosome 2B at a location that was not associated with any chlorophyll QTL effect. Independent genetic control of booting and grain filling stage heat tolerance concurs with the findings of Wardlaw et al. (1989), in which the most tolerant cultivars at the former stage were among the least tolerant at the latter stage. Different developmental stages are sensitive to the floret fertility and grain size responses to heat (booting or anthesis vs. early grain filling, respectively). Therefore, different genes/mechanisms control heat responses of grain weight and number, and breeders need to select the two types of tolerance separately, whether by using markers or phenotyping.

Loci for male sterility that is dependent on particular conditions of day length and/or temperature have been of interest in rice and wheat due to their potential utility in hybrid breeding (reviewed by Fan and Zhang, 2018 and Selva et al., 2020). The *Wtms1*, *TmsBS20T*, and *Wptms2* loci on wheat chromosome 2B are examples. These appeared to be separated from *WtmsDW* on the basis of map position (**Figure 3**). However, further work will be needed to confirm the separate location of *WtmsDW*,

due to potential problems associated with comparing maps from different studies. *Wtms1* and *TmsBS20T* differ from *WtmsDW* in expressing sterility under cold conditions (< 10°C, during spikelet differentiation stage; Xing et al., 2003; Li et al., 2006; Ru et al., 2015) rather than heat, which further supports the proposal that they are different to *WtmsDW*. It was not resolved whether it was long days and/or high temperatures during head development that was required for sterility expression by *Wptms2* (Guo et al., 2006a,b). Sterility sources considered for hybrid breeding are often described as mutant variants (e.g., *Wtms1* and *TmsBS20T*; Xing et al., 2003; Ru et al., 2015) and provide near-total sterility. By contrast, *WtmsDW* is defined by variation between two wheat cultivars. Lines carrying the intolerance allele maintained some fertility under rather severe heat conditions (**Figure 4**). These initial data suggest *WtmsDW* may not be suitable for use in hybrid wheat breeding.

The DH line WW28450 carried a deletion of the whole of the short arm of chromosome 2B and was heat intolerant. This suggested that tolerance from 2BS derives from positive gene function(s) (as opposed to absence/reduction of a tolerance suppressor), which is missing/reduced in the intolerance allele. However, 2BS carries additional male fertility loci, including *Wptms2* that conditions sterility under high temperatures and/or long days (**Figure 3**; Guo et al., 2006a), so the implications for *WtmsDW* are not entirely clear.

Effect of Heat Stress and Heat Stress Timing on Spikelet Fertility

In wheat, floret differentiation usually begins in the middle of the spike and proceeds outwards, and within each spikelet, it begins at the basal most florets and proceeds upwards. The developmental timing differs by up to 5 days across the whole spike (Evans et al., 1972; Tashiro and Wardlaw, 1990b). In the intolerant DH lines carrying the Drysdale *WtmsDW* allele, the various floret positions peaked in heat susceptibility (**Figure 2**) at times that were broadly consistent with their expected sequence of development. The florets that were the first to develop (i.e., no. 1 and 2 in spikelets from the middle of the spike), peaked in sensitivity at around 5.5 cm AD (**Figure 2A**), which was 8.6 days (heat) or 11.7 days (control) before anthesis in these florets (based on average timing for 3 and 9 cm AD). In these floret positions and at this AD, anthers were found to be at the meiosis to young microspore stage, depending on the wheat cultivar (Browne et al., 2018; Fernández-Gómez et al., 2020). Our results were in general agreement with other wheat studies that showed a peak in sensitivity to the floret sterility effects of heat stress at 18 days before mid-anthesis (Barber et al., 2017), 6–8 days before anthesis (Prasad and Djanaguiraman, 2014), during meiosis to young microspore stages of pollen development (Saini and Aspinall, 1982), or during premeiotic interphase to late leptotene stage of meiosis (Draeger and Moore, 2017). A similar stage is most sensitive to the fertility responses to drought stress (young microspore stage, at AD 4–6 cm; Ji et al., 2010), suggesting that the same developmental process may be sensitive to both heat and drought. It should be emphasized that our heat treatments did not cover the period 2–3 days before anthesis,

which is another stage where wheat can be sensitive to floret fertility effects of heat (Tashiro and Wardlaw, 1990a; Prasad and Djanaguiraman, 2014; Barber et al., 2017).

In floret positions >2, heat stress caused most sterility when it was applied at the latest stages of stem development (6.5 and 9 cm AD), which was consistent with the fact that these florets differentiate after floret positions 1 and 2 (Figure 1A). Unexpectedly, heat exposure at the earliest stage of stem development (1.6 and 3.0 cm AD) enhanced fertility in floret positions >2, particularly in the top third of the spike (Figure 1A). The biological basis for this enhanced fertility is unknown. However, for practical purposes, it had little impact on yield, because even under heat conditions, these floret positions contributed only a small proportion (< 5%) of the grains.

Floret Fertility Effects Potentially Arising by Escape

In studies of fertility responses to abiotic stresses, AD has often been used as a measure of stem development to time treatments (e.g., Ji et al., 2010; Jagadish et al., 2014). However, any genetic effect that alters the relationship between AD and spike developmental stage could potentially give escape artifacts. The two major dwarfing loci *Rht-B1* and *Rht-D1* appeared to have had such an effect in the Drysdale × Waagan DH population. This population segregated at both loci (Shirdelmoghanloo et al., 2016), and hence included lines that were double-dwarf, semi-dwarf, or tall (carrying dwarfing alleles at both, one or neither of these loci, respectively). Peak sensitivity to the effects of heat on floret sterility occurred at shorter ADs in the double-dwarfs than in the other two classes (Supplementary Figure 4), presumably because double-dwarfs had shorter AD at each of the corresponding pollen developmental stages, owing to their overall shorter AD. Similarly, tall alleles at both *Rht* loci (*Rht-B1a* and *Rht-D1a*) increased the time interval between reaching target AD and anthesis (Supplementary Table 2), indicating that the spike had further to develop at the target AD in the tall genotypes, as compared with the short genotypes. The *Rht-D1* locus also showed QTL effects for floret fertility responses, with the Waagan (tall) allele providing “tolerance” for stems that were heat exposed at the shortest AD (1.6 cm) and the Drysdale (short) allele providing tolerance for stems exposed at the longest AD (9 cm) (Supplementary Table 2). This was consistent with a potential escape mechanism in which spikes were relatively immature at a given AD in plants with potential to be tall (favoring heat exposure prior to the sensitive floret stage), and relatively mature at a given AD in plants with potential to be short (favoring heat exposure after the sensitive stage). This interpretation is supported by the results of Browne et al. (2018) and Fernández-Gómez et al. (2020). They found pollen development stage to be further advanced at a given AD in the semi-dwarf wheats Cranbrook, Young, and Wyalkatchem as compared with the tall cultivars Cadenza and Halberd (*Rht* genotype of these cultivars based on Pearce et al. (2011); and our own unpublished data). For example, an AD of 5.5 cm corresponded to meiosis in the tall cultivars but to post-meiosis (young microspore stage) in the semi-dwarf cultivars, in the most advanced florets of the spike. In the Drysdale × Waagan population, there was a floret fertility

response QTL effect on chromosome 4B (*QTL18*), but this was located ~48 cM from *Rht-B1*. Why there was no fertility response (potential escape) effect observed at *Rht-B1* is unknown.

Other studies found that *Rht* genes influenced responses of floret fertility to heat at booting, although there were inconsistencies, with reported effects ranging from positive to negative, or neutral (Alghabari et al., 2014 and references therein; Barber et al., 2017). The aforementioned issues around staging may at least partly explain these inconsistencies. However, Barber et al. (2017) reported that *Rht-D1* and the flowering time locus *Ppd-D1* affected responses to heat at around anthesis, which is not readily explained by staging artifacts, since anthesis in wheat occurs out of the boot and can usually be observed directly. On this basis, dwarfing/flowering time genes may be capable of genuinely affecting heat tolerance, at least at the anthesis stage.

We established that *WtmsDW* is a genuine heat tolerance locus, as it was not associated with any other trait that indicated the possibility of escape. Awn length at maturity was the only other trait ascribed to this QTL region (Supplementary Table 2). The flowering time locus *Ppd-B1* was mapped 7.2 cM distal of *WtmsDW* using a KASP marker in the *Ppd-B1* gene sequence (Figure 3; Supplementary Figure 1). There was no evidence that the Drysdale × Waagan DH population segregated for functional differences at *Ppd-B1*, at least under these growth conditions, because there were no flowering time QTL effects detected at this location. The closest flowering time effect was located at 26.1 cM (current study) or 5.4 cM (Shirdelmoghanloo et al., 2016) on chromosome 2B (0.7 to 1.5 day effect, with the Drysdale allele conferring lateness), which is well above both *Ppd-B1* (at 74.3 cM) and *WtmsDW* (at 81.5 cM). We have also separated *WtmsDW* from the *Ppd-B1* marker in additional fine mapping work (manuscript in preparation), confirming the independent nature of these loci. There was also no indication that DH lines with the *WtmsDW* tolerance (Waagan) allele had a sensitive stage peaking at just beyond the AD range when treatments were applied (i.e., might have escaped); these lines maintained high levels of fertility for heat treatments across the AD range of 1 to 12 cm (Supplementary Figures 5C,D).

Some fertility *per se* QTL effects were detected at *WtmsDW* under control conditions, but compared with *per se* effects under heat, these were ~3–10 times weaker in additive effect for comparable floret types and detected less frequently (Supplementary Table 2). The occasional effects observed for control conditions may have been due to a moderately hot day of 27.2°C experienced in the greenhouse during booting. Expression of *WtmsDW* floret fertility effects therefore seemed largely limited to heat stress conditions.

Evidence more or less suggested the remaining four (weaker) fertility response QTL could have been due to escape (Supplementary Table 2). This evidence seemed weakest for *QTL43* and *QTL32*. *QTL43* (43.2 to 67 cM on linkage group 7A2) showed fertility responses and fertility *per se* effects under heat stress and had no developmental effects co-locating with it (Supplementary Table 2). However, height effects mapped nearby (at 24.3 to 33.7 cM in the *QTL28* region). The floret fertility response effect at *QTL32* (66.1 cM on chromosome 1B) also mapped close to floret fertility *per se* effects under heat (at 26.2 cM in the *QTL31* region). However, the *QTL32* region also

influenced the degree of awn emergence in tillers on the day that AD on main stems reached 9 cm, implying that it may have affected the relationship between spike stage and AD. Additional work would be needed to validate *QTL43* and *QTL32* fertility tolerance effects, including ruling out potential escape artifacts. However, given the relatively weak effects of these loci (**Table 2**), they may not be worth the effort.

The fertility tolerance effects at *QTL18* and *QTL39* were much more likely to be due to escape. *QTL18* on chromosome 4B affected multiple aspects of development (**Supplementary Table 2**; Shirdelmoghanloo et al., 2016). These included flowering time, time to reach target AD, and AD at maturity, which were strong indicators of an escape artifact. Likewise, *QTL39* was associated with plant height.

Heat-Response QTL for Other Traits

Elevated temperature cannot only accelerate organ expansion but also shorten phases of development and enhance senescence (Parent et al., 2010; Asseng et al., 2011; Hunt et al., 2018). Heat stress decreased AD and plant height at maturity (**Figure 1**), perhaps by truncating the later stage of growth of these organs. For these two traits, QTL for tolerance genetic effects were observed at six loci, including *Rht-B1* and *Rht-D1* (**Table 3**). However (except at *QTL34*), the tolerance alleles were associated with lower *per se* value under control (and heat), consistent with escape, i.e., genotypes with potential for shorter height or AD had completed more of their potential growth at the time of reaching the target AD for heat treatment, and hence had less opportunity to be affected by the heat stress. *QTL34* on chromosome 2A showed a tolerance effect for height but no *per se* height effect, raising the possibility that it may have been a genuine tolerance effect.

Heat stress enhanced spike length at maturity by increasing rachis internode length but decreased awn length (as measured from the last glume). *QTL9* and *QTL29* affected heat responses of spike length and awn length, respectively, with positive alleles for heat tolerance effects (conditioning greater positive response and smaller negative response, respectively) also being positive for greater length *per se* of these organs. Elevated temperature therefore appeared to enhance further the tendency of these alleles to promote organ length, perhaps because these organs would have been in a phase of rapid growth during the treatments (based on comparing target ADs with the data of Browne et al., 2018). The *QTL41* awn length response effect was not associated with any other trait but was relatively weak.

Typically, a small number of spikelets at the bottom of the wheat spike are relatively underdeveloped and do not set grain. Five loci (*QTL5.2*, 18, 25, and both *Rht* loci) affected the tendency of heat treatment to convert such spikelets to a “developed” state (defined in this study as awns longer than half the length of those from the middle of the spike; **Table 3**). Except at *QTL5.2*, the alleles that favored this conversion conferred higher *per se* numbers of underdeveloped spikelets, probably reflecting the fact that genotypes with the potential to have higher numbers of underdeveloped spikelets had more such spikelets to convert. These loci also affected plant height and/or AD at maturity, with tall alleles being associated with a lower *per se* number of underdeveloped spikelets, suggesting a physiological

link between plant height and the ability of these basal spikelets to develop further.

QTL29 affected the time interval from target AD to anthesis, without affecting AD at maturity or flowering time. Heat magnified the genetic effect, as the Drysdale allele was positive for both the *per se* trait in control and heat tolerance effect. It was surprising that *QTL29* showed no fertility response (escape) effect, as the findings implied it affected the relationship between spike developmental stage and AD.

QTL30 on chromosome 1A and *QTL33* on chromosome 1B only affected heat responses of flowering time. These loci may represent functionally orthologous genes, since these two chromosomes are related (orthologous), and the corresponding positions of the loci were only 25–30 cM apart on the respective maps, based on BLAST searches with the peak markers (data not shown).

CONCLUSION

The *WtmsDW* locus on the short arm of wheat chromosome 2B defines a major natural variation for responses of male fertility to heat stress at booting, suggesting *WtmsDW*-linked markers may have substantial value in heat tolerance breeding. In lines carrying the *WtmsDW* intolerance allele, peaks in sensitivity of the various different floret positions were defined in relation to auricle distance. Mapping of height and flowering time traits proved very useful in identifying when floret fertility response QTL were likely to be due to escape artifacts. These insights should be valuable for guiding future efforts to screen for booting stage heat tolerance in wheat.

AUTHOR'S NOTE

This article is dedicated to the memory of Colin F. Jenner.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

ME and NC conceived the study and drafted the manuscript. JT generated the experimental design. IM-S conducted the phenotypic analyses. ME conducted the QTL analyses. ME, IL, LE, and NC performed the experimental work. All authors read and revised the manuscript.

FUNDING

ME was a recipient of a University of Adelaide ASI scholarship. This work was supported by the Grains Research and Development Corporation (GRDC) Project UA00147, the Australian Research Council (ARC) project LP150100694,

NSW-DPI, the NSW BioFirst Initiative Project 2 SAP E2985, and the Australian Centre for Plant Functional Genomics, funded by the ARC, the GRDC and the South Australian Government.

ACKNOWLEDGMENTS

We acknowledge the use of the facilities and scientific and technical assistance of the Australian Plant Phenomics Facility, which was supported by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS). QTL

mapping data previously appeared in the Ph.D. thesis of Erena (2018). We thank Hamid Shirdelmoghanloo for guidance with QTL mapping, Kerry Taylor for assistance with RIL development, and Margaret Pallotta for critically reviewing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.635397/full#supplementary-material>

REFERENCES

- Ababaei, B., and Chenu, K. (2020). Heat shocks increasingly impede grain filling but have little effect on grain setting across the Australian wheatbelt. *Agric. For. Meteorol.* 284, 107889. doi: 10.1016/j.agrformet.2019.107889
- Alghabari, F., Lukac, M., Jones, H. E., and Gooding, M. J. (2014). Effect of *Rht* alleles on the tolerance of wheat grain set to high temperature and drought stress during booting and anthesis. *J. Agron. Crops Sci.* 200, 36–45. doi: 10.1111/jac.12038
- Alonge, M., Shumate, A., Puiu, D., Zimin, A., and Salzberg, S. L. (2020). Chromosome-scale assembly of the bread wheat genome, *Triticum aestivum*, reveals over 5700 new genes. *bioRxiv* [Preprint]. doi: 10.1101/2020.04.06.028746
- Asseng, S., Ewert, F., Martre, P., Rötter, R. P., Lobell, D. B., Cammarano, D., et al. (2015). Rising temperatures reduce global wheat production. *Nat. Clim. Change* 5, 143–147. doi: 10.1038/NCLIMATE2470
- Asseng, S., Foster, I., and Turner, N. C. (2011). The impact of temperature variability on wheat yields. *Glob. Change Biol.* 17, 997–1012. doi: 10.1111/j.1365-2486.2010.02262.x
- Barber, H. M., Lukac, M., Simmonds, J., Semenov, M. A., and Gooding, M. J. (2017). Temporally and genetically discrete periods of wheat sensitivity to high temperature. *Front. Plant Sci.* 8:51. doi: 10.3389/fpls.2017.00051
- Bhusal, N., Sarial, A. K., Sharma, P., and Sareen, S. (2017). Mapping QTLs for grain yield components in wheat under heat stress. *PLoS One* 12:e0189594. doi: 10.1371/journal.pone.0189594
- Browne, R. G., Iacuone, S., Li, S. F., Dolferus, R., and Parish, R. W. (2018). Anther morphological development and stage determination in *Triticum aestivum*. *Front. Plant Sci.* 9:228. doi: 10.3389/fpls.2018.00228
- Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., and Thompson, R. (2017). *ASReml-R Reference Manual Version 4*. Hemel Hempstead: VSN International Ltd.
- Cao, Z., Li, Y., Tang, H., Zeng, B., Tang, X., Long, Q., et al. (2020). Fine mapping of the *qHTB1-1*QTL, which confers heat tolerance at the booting stage, using an *Oryza rufipogon* Griff. introgression line. *Theor. Appl. Genet.* 133, 1161–1175. doi: 10.1007/s00122-020-03539-7
- Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., et al. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8057–8062. doi: 10.1073/pnas.1217133110
- Craufurd, P. Q., Vadez, V., Jagadish, S. V. K., Prasad, P. V. V., and Zaman-Allah, M. (2013). Crop science experiments designed to inform crop modeling. *Agric. For. Meteorol.* 170, 8–18. doi: 10.1016/j.agrformet.2011.09.003
- Draeger, T., and Moore, G. (2017). Short periods of high temperature during meiosis prevent normal meiotic progression and reduce grain number in hexaploid wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 130, 1785–1800. doi: 10.1007/s00122-017-2925-1
- El Hassouni, K., Belkadi, B., Filali-Maltouf, A., Tidiane-Sall, A., Al-Abdallat, A., Nacht, M., et al. (2019). Loci controlling adaptation to heat stress occurring at the reproductive stage in durum wheat. *Agronomy* 9, 414. doi: 10.3390/agronomy9080414
- Erena, M. (2018). *Genetic and Physiological Bases of Heat Induced floret Sterility in Wheat*. Ph.D. thesis, University of Adelaide, Adelaide, SA.
- Evans, L. T., Bingham, J., and Roskams, M. A. (1972). The pattern of grain set within ears of wheat. *Aust. J. Biol. Sci.* 25, 1–8. doi: 10.1071/BI9720001
- Fan, Y., and Zhang, Q. (2018). Genetic and molecular characterization of photoperiod and thermo-sensitive male sterility in rice. *Plant Reprod.* 31, 3–14. doi: 10.1007/s00497-017-0310-5
- Fernández-Gómez, J., Talle, B., Tidy, A. C., and Wilson, Z. A. (2020). Accurate staging of reproduction development in Cadenza wheat by non-destructive spike analysis. *J. Exp. Bot.* 71, 3475–3484. doi: 10.1093/jxb/eraa156
- Gilmour, A. R., Cullis, B. R., and Verbyla, A. P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *J. Agric. Biol. Environ. Stat.* 2, 269–293. doi: 10.2307/1400446
- Guo, J., Shi, W., Guo, J., Yue, L., Zhuang, L., Zhang, W., et al. (2020). Genome-wide association studies on heat stress tolerance during grain development in wheat (*Triticum aestivum* L.). *Research Square* [Preprint] doi: 10.21203/rs.2.0460/v1
- Guo, R.-X., Sun, D.-F., Cheng, X.-D., Rong, D.-F., and Li, C. (2006b). Inheritance of thermo-photoperiod sensitive male sterility in wheat. *Aust. J. Agric. Res.* 57, 187–192. doi: 10.1071/AR05161
- Guo, R. X., Sun, D. F., Tan, Z. B., Rong, D. F., and Li, C. D. (2006a). Two recessive genes controlling thermophotoperiod-sensitive male sterility in wheat. *Theor. Appl. Genet.* 112, 1271–1276. doi: 10.1007/s00122-006-0228-z
- Hunt, J. R., Hayman, P. T., Richards, R. A., and Passioura, J. B. (2018). Opportunities to reduce heat damage in rain-fed wheat crops based on plant breeding and agronomic management. *Field Crops Res.* 224, 126–138. doi: 10.1016/j.fcr.2018.05.012
- Jagadish, K. S. V., Craufurd, P., Shi, W., and Oane, R. (2014). A phenotypic marker for quantifying heat stress impact during microsporogenesis in rice (*Oryza sativa* L.). *Funct. Plant Biol.* 41, 48–55. doi: 10.1071/FP13086
- Ji, X., Shiran, B., Wan, J., Lewis, D. C., Jenkins, C. L. D., Condon, A. G., et al. (2010). Importance of pre-anthesis anther sink strength for maintenance of grain number during reproductive stage water stress in wheat. *Plant Cell Environ.* 33, 926–942. doi: 10.1111/j.1365-3040.2010.02130.x
- Khlaïmongkhon, S., Chakhonkaen, S., Pitngam, K., Dithab, K., Sangarwut, N., Panyawut, N., et al. (2019). Molecular markers and candidate genes for thermo-sensitive genic male sterile in rice. *Rice Sci.* 26, 147–156. doi: 10.1016/j.rsci.2018.08.006
- Kuchel, H., Williams, K., Langridge, P., Eagles, H. A., and Jefferies, S. P. (2007). Genetic dissection of grain yield in bread wheat. II. QTL-by-environment interaction. *Theor. Appl. Genet.* 115, 1015–1027. doi: 10.1007/s00122-007-0628-8
- Lemerle, D., Smith, A., Verbeek, B., Koetz, E., Lockley, P., and Martin, P. (2006). Incremental crop tolerance to weeds: a measure for selecting competitive ability in Australian wheats. *Euphytica* 149, 85–95. doi: 10.1007/s10681-005-9056-5

- Li, J., and Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 95, 221–227. doi: 10.1038/sj.hdy.6800717
- Li, Y. F., Zhao, C. P., Zhang, F. T., Sun, H., and Sun, D. F. (2006). Fertility alteration in the photo-thermo-sensitive male sterile line BS20 of wheat (*Triticum aestivum* L.). *Euphytica* 151, 207–213. doi: 10.1007/s10681-006-9141-4
- Lu, L., Liu, H., Wu, Y., and Yan, G. (2020). Development and characterization of near-isogenic lines revealing candidate genes for a major 7AL QTL responsible for heat tolerance in wheat. *Front. Plant Sci.* 11:1316. doi: 10.3389/fpls.2020.01316
- Mahjourimajd, S., Taylor, J., Sznajder, B., Timmins, A., Shahinnia, F., Rengel, Z., et al. (2016). Genetic basis for variation in wheat grain yield in response to varying nitrogen application. *PLoS One* 11:e0159374. doi: 10.1371/journal.pone.0159374
- Malosetti, M., Voltas, J., Romagosa, I., Ullrich, S. E., and van Eeuwijk, F. A. (2004). Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica* 137, 139–145. doi: 10.1023/B:EUPH.0000040511.46388.ef
- Mason, R. E., Hays, D. B., Mondal, S., Ibrahim, A. M. H., and Basnet, B. R. (2013). QTL for yield, yield components and canopy temperature depression in wheat under late sown field conditions. *Euphytica* 194, 243–259. doi: 10.1007/s10681-013-0951-x
- Mason, R. E., Mondal, S., Beecher, F. W., and Hays, D. B. (2011). Genetic loci linking improved heat tolerance in wheat (*Triticum aestivum* L.) to lower leaf and spike temperatures under controlled conditions. *Euphytica* 180, 181–194. doi: 10.1007/s10681-011-0349-6
- Mason, R. E., Mondal, S., Beecher, F. W., Pacheco, A., Jampala, B., Ibrahim, A. M. H., et al. (2010). QTL associated with heat susceptibility index in wheat (*Triticum aestivum* L.) under short-term reproductive stage heat stress. *Euphytica* 174, 423–436. doi: 10.1007/s10681-010-0151-x
- McDonald, G., Bovill, W., Taylor, J., and Wheeler, R. (2015). Responses to phosphorus among wheat genotypes. *Crops Pasture Sci.* 66, 430–444. doi: 10.1071/CP14191
- Mohammadi, V., Zali, A. A., and Bihamta, M. R. (2008). Mapping QTLs for heat tolerance in wheat. *J. Agric. Sci. Technol.* 10, 261–267.
- Nubankoh, P., Wanchana, S., Saensuk, C., Ruanjaichon, V., Cheabu, S., Vanavichit, A., et al. (2020). QTL-seq reveals genomic regions associated with spikelet fertility in response to a high temperature in rice (*Oryza sativa* L.). *Plant Cell Rep.* 39, 149–162. doi: 10.1007/s00299-019-02477-z
- Pandey, G. C., Sareen, S., Siwach, P., and Tiwari, R. (2014). Molecular characterization of heat tolerance in bread wheat (*Triticum aestivum* L.) using differences in thousand-grain weights (dTGW) as a potential indirect selection criterion. *Cereal Res. Commun.* 42, 38–46. doi: 10.1556/CRC.2013.0041
- Parent, B., Turc, O., Gibon, Y., Stitt, M., and Tardieu, F. (2010). Modelling temperature-compensated physiological rates, based on the co-ordination of responses to temperature of developmental processes. *J. Exp. Bot.* 61, 2057–2069. doi: 10.1093/jxb/erq003
- Patterson, H. D., and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554. doi: 10.1093/biomet/58.3.545
- Payne, R. W. (2009). GenStat. *WIREs Comput. Stat.* 1, 255–258. doi: 10.1002/wics.32
- Pearce, S., Saville, R., Vaughan, S. P., Chandler, P. M., Wilhelm, E. P., Sparks, C. A., et al. (2011). Molecular characterization of Rht-1 dwarfing genes in hexaploid wheat. *Plant Physiol.* 157, 1820–1831. doi: 10.1104/pp.111.183657
- Pinto, R. S., Reynolds, M. P., Mathews, K. L., McIntyre, C. L., Olivares-Villegas, J.-J., and Chapman, S. C. (2010). Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. *Theor. Appl. Genet.* 121, 1001–1021. doi: 10.1007/s00122-010-1351-4
- Pradhan, S., Ali Babar, M., Robbins, K., Bai, G., Mason, R. E., Khan, J., et al. (2019). Understanding the genetic basis of spike fertility to improve grain number, harvest index, and grain yield in wheat under high temperature stress environments. *Front. Plant Sci.* 10:1481. doi: 10.3389/fpls.2019.01481
- Prasad, P. V. V., and Djanaguiraman, M. (2014). Response of floret fertility and individual grain weight of wheat to high temperature stress: sensitive stages and thresholds for temperature and duration. *Funct. Plant Biol.* 41, 1261–1269. doi: 10.1071/FP14061
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Romagosa, I., Han, F., Ullrich, S. E., Hayes, P. M., and Wesenberg, D. M. (1999). Verification of yield QTL through realized molecular marker-assisted selection responses in a barley cross. *Mol. Breed.* 5, 143–152. doi: 10.1023/A:1009684108922
- Ru, Z.-G., Zhang, L.-P., Hu, T.-Z., Liu, H.-Y., Yang, Q.-K., Weng, M.-L., et al. (2015). Genetic analysis and chromosome mapping of a thermo-sensitive genic male sterile gene in wheat. *Euphytica* 201, 321–327. doi: 10.1007/s10681-014-1218-x
- Saini, H. S., and Aspinall, D. (1982). Abnormal sporogenesis in wheat (*Triticum aestivum* L.) induced by short periods of high temperature. *Ann. Bot.* 49, 835–846. doi: 10.1093/oxfordjournals.aob.a086310
- Selva, C., Riboni, M., Baumann, U., Würschum, T., Whitford, R., and Tucker, M. R. (2020). Hybrid breeding in wheat: how shaping floral biology can offer new perspectives. *Funct. Plant Biol.* 47, 675–694. doi: 10.1071/FP19372
- Sharma, D., Singh, R., Rane, J., Gupta, V. K., Mamrutha, H. M., and Tiwari, R. (2016). Mapping quantitative trait loci associated with grain filling duration and grain number under terminal heat stress in bread wheat (*Triticum aestivum* L.). *Plant Breed.* 135, 538–545. doi: 10.1111/pbr.12405
- Shirdelmoghannloo, H., Taylor, J. D., Lohraseb, I., Rabie, H., Brien, C., Timmins, A., et al. (2016). A QTL on the short arm of wheat (*Triticum aestivum* L.) chromosome 3B affects the stability of grain weight in plants exposed to a brief heat shock early in grain filling. *BMC Plant Biol.* 16:100. doi: 10.1186/s12870-016-0784-6
- Stone, P. J., and Nicolas, M. E. (1995). Effect of timing of heat stress during grain filling on two wheat varieties differing in heat tolerance. I. Grain growth. *Aust. J. Plant Physiol.* 22, 927–934. doi: 10.1071/PP9950927
- Stone, P. J., and Nicolas, M. E. (1998). Comparison of sudden heat stress with gradual exposure to high temperature during grain filling in two wheat varieties differing in heat tolerance. II. Fractional protein accumulation. *Aust. J. Plant Physiol.* 25, 1–11. doi: 10.1071/PP96103
- Tashiro, T., and Wardlaw, I. F. (1990a). The effect of high temperature at different stages of ripening on grain set, grain weight and grain dimensions in the semi-dwarf wheat 'Banks'. *Ann. Bot.* 65, 51–61. doi: 10.1093/oxfordjournals.aob.a087908
- Tashiro, T., and Wardlaw, I. F. (1990b). The response to high temperature shock and humidity changes prior to and during the early stages of grain development in wheat. *Aust. J. Plant Physiol.* 17, 551–561. doi: 10.1071/PP9900551
- VSN International. (2020). *Genstat for Windows*, 21st Edn. Hemel Hempstead: VSN International.
- Wardlaw, I. F., Dawson, I. A., Munibi, P., and Fewster, R. (1989). The tolerance of wheat to high temperatures during reproductive growth. I. Survey procedures and general response patterns. *Aust. J. Agric. Res.* 40, 1–13. doi: 10.1071/AR9890001
- Watson-Haigh, N. S., Suchecki, R., Kalashyan, E., Garcia, M., and Baumann, U. (2018). DAWN: a resource for yielding insights into the diversity among wheat genomes. *BMC Genomics* 19:941. doi: 10.1186/s12864-018-5228-2
- Wrigley, C. W. (2007). "Mitigating the damaging effects of growth and storage conditions on grain quality, in Wheat Production in stressed environments," in *Proceedings of the 7th International Wheat Conference, 27 November - 2 December 2005, Mar del Plata, Argentina*, eds H. T. Buck, J. E. Nisi, and N. Salomón (Dordrecht: Springer), 425–439.
- Xing, Q. H., Ru, Z. G., Zhou, C. J., Xue, X., Liang, C. Y., Yang, D. E., et al. (2003). Genetic analysis and molecular mapping of thermo-sensitive genic male-sterile gene (*wtms1*) in wheat. *Theor. Appl. Genet.* 107, 1500–1504. doi: 10.1007/s00122-003-1385-y
- Yu, J., Han, J., Kim, Y.-J., Song, M., Yang, Z., He, Y., et al. (2017). Two rice receptor-like kinases maintain male fertility under changing temperatures. *Proc. Natl. Acad. Sci. U.S.A.* 114, 12327–12332. doi: 10.1073/pnas.1705189114

- Zadoks, J. C., Chang, T. T., and Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Res.* 14, 415–421. doi: 10.1111/j.1365-3180.1974.tb01084.x
- Zheng, B. S., Gouis, J. L., Leflon, M., Rong, W. Y., Laperche, A., and Brancourt-Hulmel, M. (2010). Using probe genotypes to dissect QTL $\times$ environment interactions for grain yield components in winter wheat. *Theor. Appl. Genet.* 121, 1501–1517. doi: 10.1007/s00122-010-1406-6
- Zhu, S., Huang, R., Wai, H. P., Xiong, H., Shen, X., He, H., et al. (2017). Mapping quantitative trait loci for heat tolerance at the booting stage using chromosomal segment substitution lines in rice. *Physiol. Mol. Biol. Plants* 23, 817–825. doi: 10.1007/s12298-017-0465-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Erena, Lohraseb, Munoz-Santa, Taylor, Emebiri and Collins. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Molecular Cytogenetic and Agronomic Characterization of the Similarities and Differences Between Wheat–*Leymus mollis* Trin. and Wheat–*Psathyrostachys huashanica* Keng 3Ns (3D) Substitution Lines

OPEN ACCESS

Edited by:

Shuyu Liu,
Texas A&M University System,
United States

Reviewed by:

Cheng Liu,
Crop Research Institute, Shandong
Academy of Agricultural Sciences,
China
Bernardo Ordas,
Consejo Superior de Investigaciones
Científicas (CSIC), Spain

*Correspondence:

Xinhong Chen
cxh2089@126.com
Jixin Zhao
zhjx881@163.com

[†] These authors contributed equally to
this work

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 22 December 2020

Accepted: 23 February 2021

Published: 31 March 2021

Citation:

Li J, Li J, Cheng X, Zhao L,
Yang Z, Wu J, Yang Q, Chen X and
Zhao J (2021) Molecular Cytogenetic
and Agronomic Characterization
of the Similarities and Differences
Between Wheat–*Leymus mollis* Trin.
and Wheat–*Psathyrostachys*
huashanica Keng 3Ns (3D)
Substitution Lines.
Front. Plant Sci. 12:644896.
doi: 10.3389/fpls.2021.644896

Jiachuang Li[†], Jiaojiao Li[†], Xueni Cheng², Li Zhao¹, Zujun Yang³, Jun Wu¹,
Qunhui Yang¹, Xinhong Chen^{1*} and Jixin Zhao^{4*}

¹ Shaanxi Key Laboratory of Plant Genetic Engineering Breeding, College of Agronomy, Northwest A&F University, Xianyang, China, ² College of Life Sciences, Northwest A&F University, Xianyang, China, ³ School of Life Sciences and Technology, University of Electronic Science and Technology of China, Chengdu, China, ⁴ College of Agronomy, Northwest A&F University, Xianyang, China

Psathyrostachys huashanica Keng ($2n = 2x = 14$, NsNs) and *Leymus mollis* Trin. ($2n = 4x = 28$, NsNsXmXm) are valuable resources for wheat breeding improvement as they share the Ns genome, which contains diverse resistance genes. To explore the behaviors and traits of Ns chromosomes from the two species in wheat background, a series of wheat–*P. huashanica* and wheat–*L. mollis* substitution lines were developed. In the present study, line DH109 (F₇ progeny of wheat–*P. huashanica* heptaploid line H8911 × durum wheat Trs-372) and line DM131 (F₈ progeny of wheat–*L. mollis* octoploid line M842 × durum wheat Trs-372) were selected. Cytological observation combined with genomic *in situ* hybridization experiments showed that DH109 and DM131 each had 20 pairs of wheat chromosomes plus a pair of alien chromosomes (Ns chromosome), and the pair of alien chromosomes showed stable inheritance. Multiple molecular markers and wheat 55K SNP array demonstrated that a pair of wheat 3D chromosome in DH109 and in DM131 was substituted by a pair of *P. huashanica* 3Ns chromosome and a pair of *L. mollis* 3Ns chromosome, respectively. Fluorescence *in situ* hybridization (FISH) analysis confirmed that wheat 3D chromosomes were absent from DH109 and DM131, and chromosomal FISH karyotypes of wheat 3D, *P. huashanica* 3Ns, and *L. mollis* 3Ns were different. Moreover, the two lines had many differences in agronomic traits. Comparing with their wheat parents, DH109 expressed superior resistance to powdery mildew and fusarium head blight, whereas DM131 had powdery mildew resistance, longer spike, and more tiller number. Therefore, Ns genome from *P. huashanica* and *L. mollis* might have some different effects. The two novel wheat–alien substitution lines provide new ideas and resources for disease resistance and high-yield breeding on further utilization of 3Ns chromosomes of *P. huashanica* or *L. mollis*.

Keywords: wheat, *Psathyrostachys huashanica*, *Leymus mollis*, Ns genome chromosome, substitution lines

INTRODUCTION

As one of the three major cereals in the world, common wheat (*Triticum aestivum* L., $2n = 6x = 42$, AABBDD) makes tremendous contributions to the development of human civilizations. The origin of common wheat was the result of natural distant hybridization between *Triticum* and *Aegilops* (Chantret et al., 2005). It is widely believed that wild relative species of wheat possess numerous excellent traits, i.e., disease resistance, stress tolerance, and high productivity, which are what common wheat needs. Because Barelle did the first artificial interspecific hybridization in wheat in the early 19th century and found that offspring had improved adaption to different environments (Ciferri, 1955), interspecific and intergeneric hybridization of common wheat were always important directions for breeders. So far, nearly 90 species from 14 relative genera of wheat have successfully crossed with common wheat, and a series of wheat-related species germplasm resources with outstanding agronomic traits have been developed (Pauk, 2016). For example, wheat-*Secale cereale* 1BL/1RS translocation, 1B (R) substitution, and 4R addition lines had high productivity and resistance to stripe rust (Rabinovich, 1998; An et al., 2013); wheat-*Haynaldia villosa* 6VS/6AL and 4VS/4DL translocation lines conferred resistance to wheat powdery mildew and streak mosaic virus, respectively (Chen et al., 1995; Zhang et al., 2005); wheat-*Thinopyrum ponticum* 7Jst (7B) and 1st (1B) + 4St-4Jst (4B) substitution lines exhibited resistance to wheat stripe rust (Li et al., 2015; Zhu et al., 2017); and wheat-*Agropyron cristatum* 6P addition and 2P translocation lines had high resistance to wheat powdery mildew (Han et al., 2014; Jiang et al., 2018).

Psathyrostachys huashanica Keng ($2n = 2x = 14$, NsNs) belongs to *Psathyrostachys* Nevski, which is diploid perennial plant containing approximately 10 species possessing Ns genome (Baden, 1991; Wang et al., 1994). *P. huashanica* occurred only on the stony slopes of Huashan Mountains, Shaanxi Province, China, and owned numerous excellent traits, such as resistance to wheat disease (rust, take-all, scab, and powdery mildew), tolerance to abiotic stress (salinity, alkalinity, and cold), and early maturation (Baden, 1991; Jing et al., 1999; Wang and Shang, 2000; Song et al., 2013). The first distant hybridization between common wheat and *P. huashanica* was conducted successfully in our laboratory by Chen (1991) using embryos culture method. The F₁ hybrid H811 ($2n = 6x = 28$, ABDNs) as female parent backcrossed with common wheat line 7,182 to generate heptaploid hybrid H8911 ($2n = 7x = 49$, AABBDDNs), which was used to self-cross or cross with other wheat cultivars to develop several wheat-*P. huashanica*-derived lines.

Leymus mollis Trin. ($2n = 4x = 28$, NsNsXmXm) is a cross-pollinated heterotetraploid perennial species of *Leymus* Hochst., and it grows mostly on the coastal beaches. *L. mollis* was regarded as a suitable exogenous germplasm for wheat improvement because of its outstanding traits including long spikes, full-spikelet, strong stems, immunity to diseases caused by bacteria and fungi, and high tolerance to saline-alkali soil

(Mujeeb-Kazi and Rodriguez, 1981; Kishii et al., 2003). Distant hybridization between *Leymus* and common wheat could date back to late 1960s (Tsitsis, 1965). Later, the wheat-*L. mollis* octoploid derivative line M842-12 ($2n = 8x = 56$, AABBDDNsNs) and M842-13 ($2n = 8x = 56$, AABBDDXmXm) were created by Fu et al. (1993) via embryo rescue and colchicine treatment. Subsequently, a series of wheat-*L. mollis*-derived germplasms were obtained mainly by using octoploid *Tritileymus* M842-12 to cross with wheat cultivars.

The original studies suggested that the genome of *Leymus* was from two genera: *Thinopyrum* Löve (E genome) and *Psathyrostachys* Nevski (Ns genome) based on its rhizomatous growth habit and saline habitat (Dewey, 1984; Löve, 1984). The existence of Ns genome has been verified using cytogenetic molecular methods (Wang and Hsiao, 1984; Sun et al., 1995). But, Zhang and Dvořák (1991) raised question about the presence of E genome, and they were supported by subsequent studies from other researchers (Ørgaard and Heslop-Harrison, 1994; Wang and Jensen, 1994). As the other genome was unknown, the genome of *Leymus* was assigned as NsNsXmXm (Xm meaning the unknown genome) (Wang et al., 1994). However, up to now, which species of *Psathyrostachys* Nevski donates the Ns genome is still unclear.

In the present study, two novel wheat-alien-derived lines were developed to judge whether *P. huashanica* was a donor of Ns genome to *Leymus* and whether Ns genome chromosomes from the same homoeology but different genera would express the same agronomic traits in wheat background. The objectives of the research were to (a) develop wheat-*P. huashanica* and wheat-*L. mollis* substitution lines, (b) identify inherent stability and homoeologous group of alien chromosomes in wheat background, and (c) investigate agronomic and morphologic traits of two lines.

MATERIALS AND METHODS

Development of Plant Materials

The plant materials used in this study included *P. huashanica* Keng ($2n = 14$, NsNs), *L. mollis* pilger ($2n = 28$, NsNsXmXm), common wheat ($2n = 42$, AABBDD) lines 7,182, Mingxian169 (MX169), Chinese Spring (CS) and Huixianhong (HXH), *Triticum durum* ($2n = 28$, AABB) line Trs-372, wheat-*P. huashanica* disomic substitution line DH109, and wheat-*L. mollis* disomic substitution line DM131. DH109 was obtained from the F₇ progeny of wheat-*P. huashanica* heptaploid line H8911 × line Trs-372. The hexaploid hybrid H8911 ($2n = 49$, AABBDDNs) was generated from the cross between common wheat line 7,182 and *P. huashanica*. DM131 was obtained from the F₈ progeny of wheat-*L. mollis* octoploid line M842 × line Trs-372. The octoploid hybrid M842 ($2n = 56$, AABBDDNsNs) was generated from the cross between common wheat line 7,182 and *L. mollis*. MX169, CS, and HXH were susceptible controls in the disease resistance testing. All materials were deposited at the College of Agronomy, Northwest A&F University, China. Total genomic DNA was extracted using the standard CTAB method.

Cytological Observation

The roots and young spikes were sampled at appropriate stages, when the lengths of roots and panicles were 1–2 and 5–6 cm, respectively. Samples were pretreated in an ice–water bath for 24 h before transfer to Carnoy's fixative fluid I (ethanol: glacial acetic acid mixture at 3:1, vol/vol) for 24 h and finally to 70% ethanol and stored at -20°C . After treatment with 1% cellulase (Yakult, Japan) and 2% pectinase (Yakult, Japan) at 37°C for 1 h, the root tips were cleaved into signal cell to facilitate the observation of chromosomal number and morphology. Anthers were taken from the middle to both sides of the spike until target stages and the microsporocytes were stained with 1% acetocarmine before cytological observations. The slides with good split phases were dried and marked for the following experiments. These slides' preparation for fluorescence *in situ* hybridization (FISH) and genomic *in situ* hybridization (GISH) analysis was through UV crosslinking ($1,250\text{ mJ}/\text{cm}^2$, 3 min) which make chromosomes attached on slides. In this process, the root and spikes were numbered to ensure derivation from the same one seed. Twelve plants of each line were randomly selected for cytological screen and *in situ* hybridization analysis for 5 consecutive years.

GISH Analysis

In GISH experiment, genomic DNA from alien donor *P. huashanica* and *L. mollis* labeled with DIG-11-dUTP were used as probes for GISH analysis. After more than 16-h hybridization in a dark and moist box at 37°C , anti-digoxigenin fluorescein kit (Roche, Germany) was used to visualize the combinative zone of probes, and Vectashield H-1300 (VECTOR, United States) was used to counterstain the chromosomes. Detailed procedures could be seen in an article by Zhao et al. (2013). Fluorescent signals were observed with a microscope (ZEISS Imager M2, Germany) and imaged (ZEISS ICC5, Germany).

DNA Marker Analysis

Two hundred six pairs of simple sequence repeat (SSR) primers from each wheat chromosome were selected to determine the chromosomal composition of DH109 and DM131. SSR markers included GWM series developed by Röder et al. (1998), GDM series developed by Pestsova et al. (2000), and CFA and CFD series developed by Sourdille et al. (2004). One hundred twenty-four pairs of expressed sequence tag–sequence tag site (EST-STS) primers¹ distributed among seven wheat homoeologous groups with corresponding chromosomes of *Ns* genome were employed to determine the homoeology of the introduced alien chromosomes in two lines. In addition, nine pairs of *P. huashanica* *Ns* genome–specific sequence characterized amplified region (SCAR) markers (Chen et al., 2010; Wang et al., 2014; Su et al., 2015) located in the 1Ns, 3Ns, and 5Ns chromosomes were used for additional chromosomes verification.

The products of EST-STS and SSR markers experiments were electrophoresed on 8% non-denaturing polyacrylamide gel (constant voltage 165 V, 2.5 h) and stained with alkaline-silver

method. The products of SCAR markers were separated on 1% agarose gels (150 V, 0.5 h) and observed using BIO-RAD chemiDoc XRS + (ImageLab system, United States). The specific and parallel bands amplified from the two lines and alien parents by EST-STS markers were sequenced by Sangon, China, and aligned using DNAMAN V6.0.3 (Lynnon Biosoft, United States) and BLAST tool on NCBI² and URG1³.

FISH and Sequential GISH

In FISH experiment, a pair of fluorescent-modified probes comprising oligo-pSc119.2 (6-FAM-5') and oligo-pTa535-1 (TAMRA-5') (Danilova et al., 2012; Tang et al., 2014) was used to identify the chromosomal compositions of DH109 and DM131. Homoeologous group of each wheat chromosome could be distinguished according to fluorescent spots spread on chromosomal arm. The chromosomal FISH signals of the two materials were compared with karyotype of Chinese Spring and Mianyang 11 provided by Tang et al. (2014). Oligo-primers were dissolved in $1 \times \text{TE}$ solution to $20\text{ ng}/\mu\text{L}$. Each slide with mixture ($3\text{ }\mu\text{L}$ pSc119.2, $2\text{ }\mu\text{L}$ pTa535-1, and $5\text{ }\mu\text{L}$ $1 \times \text{TE}$) was put in a moisturizing black box at 55°C for more than 3 h. Then, slides were immersed in a $2 \times \text{SSC}$ solution to make coverslip slip. Vectashield H-1200 (VECTOR, United States) was used to counterstain the chromosomes. For the sequential GISH, the slides photographed were soaked in 75% alcohol for 5 min and exposed to light for 24 h. The protocol of sequential GISH was the same as provided in GISH analysis. Fluorescent signals were observed with a microscope (ZEISS Imager M2, Germany) and imaged (ZEISS ICC5, Germany).

Wheat 55K SNP Array Analysis

Purifying genomic DNA of materials was hybridized to wheat 55K SNP genotyping arrays, and Illumina Bead Array technology was used for scanning in China Golden Marker Biotechnology Company (Beijing, China). The wheat 55K SNP array contained 49,078 SNPs, which were distributed across 21 pairs of wheat chromosomes. The total valid number of markers divided by the marker number that had the same genotype in a chromosome between two lines was calculated as the percentage of the same markers on each chromosome. Excel 2016 (Microsoft, United States) was used for statistics and analysis of data. Row picture and chromosome map used SigmaPlot V12.5 (SYSTAT software, Inc., United States) and MapChart V2.32 (Wageningen University & Research, Netherlands), respectively.

Evaluation of Disease Resistance and Agronomic Traits of Materials

Resistance to wheat common diseases [stripe rust, powdery mildew, and fusarium head blight (FHB)] and agronomic traits of materials were evaluated for 3 consecutive years (2018–2020) in Yangling, China. In the field condition, the materials were arranged separately in a completely randomized block design, and each material had two rows with 12-cm interval of each plant. In the incubator, all materials were separated planted in one plug,

¹https://wheat.pw.usda.gov/SNP/new/pcr_primers.shtml

²<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

³<https://urgi.versailles.inrae.fr/blast/>

and the controls were in the center and four corners for better infection. Meanwhile, five plants of each material were arranged with three replications. For the evaluation of disease resistance, 12 plants of each material were investigated in the same way every year. For the evaluation of agronomic traits, the average data based on five samples and three repeats of every year to ensure that accurate results were obtained.

Six morphological traits comprising the plant height, tiller number, spike length, spikelet number, kernel number, and thousand-kernel weight were investigated. Grain quality indicators of materials, including the kernel protein content, gluten protein content, starch content, subsidence value, volume weight, dough stability time, and flour field, were tested by Perten DA 7250 NIR analyzer (Sweden). Significant analyses between different materials were conducted using the SPSS Statistics 20 software program (IBM Corp., Armonk, NY, United States).

To access the adult plant resistance to stripe rust of two lines, three different races of *Puccinia striiformis* f. sp. *tritici* (CYR32, 33, 34) were used for artificial inoculation at an appropriate period. Mixed races were smeared onto the wheat flag leaves of every experimental material after a drizzle of early spring for better infection. The susceptible cultivar was Mingxian 169, and the infection types (ITs) to stripe rust were graded according to the method provided by Ma et al. (1995). IT was rated on a scale from 0 to 4, in which 0 and 0 indicate immune and nearly immune, 1 and 2 denote high resistance and moderate resistance, and 3 and 4 indicate moderately susceptible and susceptible, respectively. Each type can be appended with “+” or “-” to indicate that it is heavier or lighter.

The evaluation of resistance to powdery mildew was conducted at the seeding stage in a growth chamber. The *Blumeria graminis* f. sp. *tritici* isolate E09 was used for inoculation when materials were at two-leaf stage, and Huixianhong was susceptible control. Pathogen spores with high activity were dusted onto the leaves, and plants were incubated at 22°C and 70% humidity for 15 days. The ITs to powdery mildew were scored using the method of Sheng (1988) on five grades, which were IT = 0 and 0 indicating immune and nearly immune, IT = 1 and 2 denoting high resistance and moderate resistance, and 3 and 4 meaning moderately susceptible and susceptible, respectively. Each type can be appended with “+” or “-” to indicate that it is heavier or lighter.

The type II resistance to FHB was evaluated at field using the method of single floret inoculation described by Bai et al. (1999; 2000). In brief, 10 randomly selected spikes at flowering period were injected 10 µL of conidial spore suspension into the floral cavity between the lemma and palea of a single floret in the middle of one spike. *Fusarium graminearum* Schwabe strain PH1 was expanded and diluted to 100 spores µL⁻¹ in mung beans liquid medium. Each inoculated spike was covered with a moist plastic bag for 2 days, and total spikelets and infected spikelets were counted at 21 days after injection (Bai, 1996). The infected grades basing on symptomatic spikelet of entire spike were 1 to 5, where 1 indicates no extension to cob and 5 means symptomatic spikelet more than three-fourths of the whole spike. Intermediate infection grades of spike were represented by 2 (less than 1/4), 3 (1/4 to 1/2), and 4 (1/2 to 3/4). Reaction index (RI, 1–5) and

infected spikelet rate (ISR, %) of materials to FHB were according to Yang et al. (1998) as follows: $RI = \Sigma (\text{number of spikes at each infected grade} \times \text{correspond infected grade}) / \text{number of total spikes}$, where RI = 1.1–2.0 denotes resistance, 2.1–3.0 denotes moderate resistance, 3.1–4.0 denotes moderately susceptible, and 4.1–5.0 denotes susceptible; $ISR = \Sigma (\text{infected spikelets} / \text{total spikelets}) / \text{number of total spikes}$.

RESULTS

Observation of Cytogenetics of DH109 and DM131

Part division phases of root tip cells (RTCs) and pollen mother cells (PMCs) were observed to clarify chromosomal numbers and pairing. The mitosis metaphase observations indicated that RTCs of line DH109 (Figure 1A) and line DM131 (Figure 1B) both had a chromosome number of 42. PMCs in meiotic metaphase I showed that the DH109 had a chromosome configuration of 21 bivalents without a trivalent or quadrivalent (Figure 1C) and so had DM131 (Figure 1D). Meiotic anaphase I of PMC could exhibit segregation of homoeologous chromosomes. In PMCs of line DH109 (Figure 1E) and line DM131 (Figure 1F), chromosomes segregated and moved to the cell poles normally at meiotic anaphase I. These results indicated that DH109 and DM131 were cytological stable lines with regular chromosome number and cell division.

GISH Analyses of DH109 and DM131

Genomic *in situ* hybridization analyses were using the whole genomic DNA from *P. huashanica* for DH109 and *L. mollis* for DM131 as the probe and common wheat parent as the block to identify number of alien chromosome in the derived line. The results of RTCs GISH analysis showed that line DH109 had two chromosomes from *P. huashanica* (Figure 2A), and line DM131 had two chromosomes from *L. mollis* (Figure 2B). GISH analysis of PMCs in meiotic metaphase I showed that a rod bivalent with yellow-green signals in DH109 (Figure 2C) and a ring bivalent with hybridization signal in DM131 (Figure 2D). In meiotic telophase II, each of the four sperms carried an alien chromosome in DH109 and DM131 according to Figures 2E,F, respectively. Mitotic correlative results demonstrated that the two lines were both disomic substitution lines in which two wheat chromosomes were substituted by Ns chromosomes, and the alien chromosomes in DH109 were from *P. huashanica* and were from *L. mollis* in DM131. PMCs' GISH analysis indicated that the two substitution lines were cytogenetically stable wheat-alien-derived lines for the alien chromosomes could pair, segregate, and inherit normally.

Multiple Molecular Markers Analysis of DH109 and DM131

Among the 206 pairs of SSR markers spread on 21 pairs of wheat chromosomes, eight pairs of markers, including *barc135*, *xcfd64*, *xcfd223*, *xgwm52*, *xgwm314*, *xgwm456*, *xgwm497*, and *xgwm645* related to wheat 3D chromosomes could amplify wheat

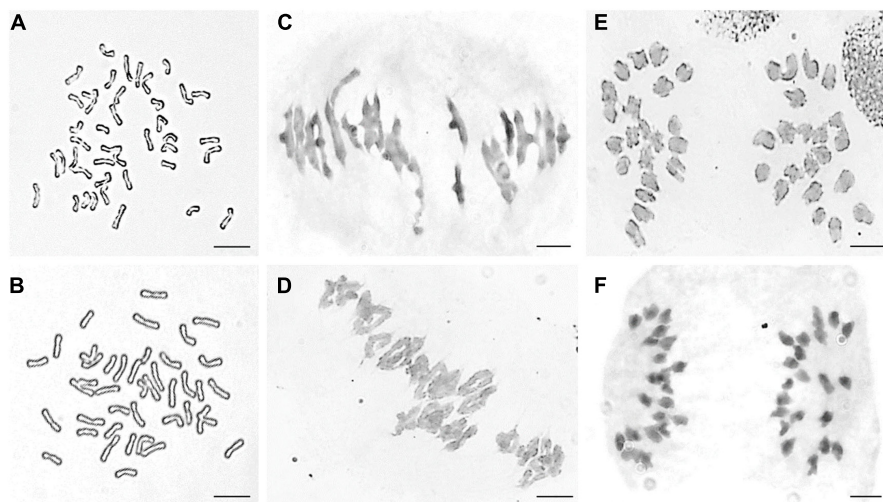


FIGURE 1 | Cytological observation of root tip cells (RTCs) and pollen mother cells (PMCs). **(A)** Mitotic metaphase of DH109, $2n = 42$. **(B)** Mitotic metaphase of DM131, $2n = 42$. **(C)** Meiotic metaphase I of DH109, $2n = 21$ II. **(D)** Meiotic metaphase I of DM131, $2n = 21$ II. **(E)** Meiotic anaphase I of DH109, $2n = 21 + 21$. **(F)** Meiotic anaphase I of DM131, $2n = 21 + 21$. Scale bar, 10 μ m.

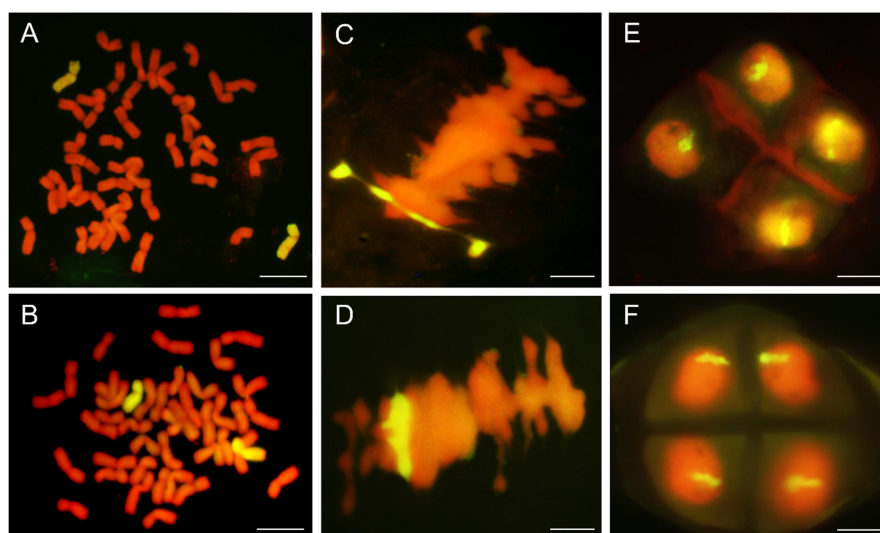
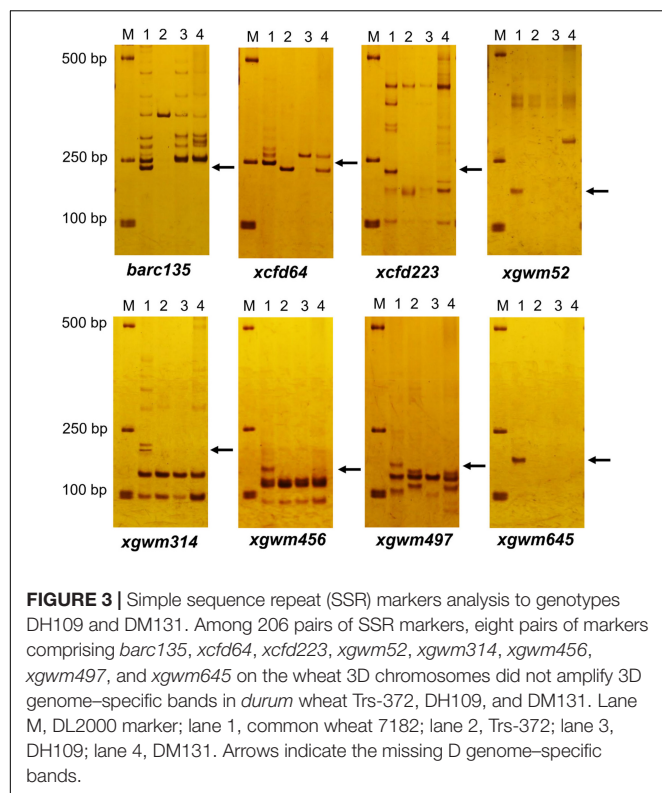


FIGURE 2 | Genomic *in situ* hybridization (GISH) analysis of line DH109 and line DM131. GISH analysis of RTCs in the mitotic metaphase. **(A)** In DH109, two chromosomes with yellow-green signals were detected as alien chromosomes from *P. huashanica*. **(B)** In DM131, two chromosomes with yellow-green signals were detected as alien chromosomes from *L. mollis*. GISH analysis of PMCs in the mitotic metaphase. **(C)** Alien chromosomes formed a rod bivalent with fluorescent signal in DH109. **(D)** Alien chromosomes formed a ring bivalent with fluorescent signal in DM131. GISH analysis of gametes in the meiosis telophase II. **(E)** Each of the four progeny cells had a fluorescent signal in DH109. **(F)** Each of the four progeny cells had a fluorescent signal in DM131. Chromosomes were counterstained with propidium iodide (red). Scale bar, 10 μ m.

D genome-specific bands in common wheat 7,182, but not in durum wheat Trs-372 and two derived lines (**Figure 3** and **Table 1**). The results indicated that both DH109 and DM131 lost their wheat 3D chromosomes.

EST-STS markers and SCAR markers could be used to identify alien chromosomal homoeology in wheat background. We selected 124 pairs of EST-STS markers that distributed among seven homoeologous groups. Ten pairs of STS primers all belonging to the 3rd homoeologous group could amplify

Ns genome-specific bands in two derived lines and alien parents, whereas these bands could not be amplified in wheat parents (**Figure 4** and **Table 1**). Among the 10 pairs of markers, three markers were applicative for *P. huashanica* Ns genome, two markers were applicative for *L. mollis* Ns genome, and four markers were universal. The polymerase chain reaction (PCR) results of SCAR markers showed that two *P. huashanica* 3Ns genome-specific SCAR markers, i.e., S3-113 and S3-125, amplified unique and clear bands in DH109 and



P. huashanica (Figure 5 and Table 1), but not in other materials. These results demonstrated that although 3Ns chromosomes were introduced into both derived lines, DH109 possessed

P. huashanica 3Ns chromosomes, and DM131 had *L. mollis* 3Ns chromosomes.

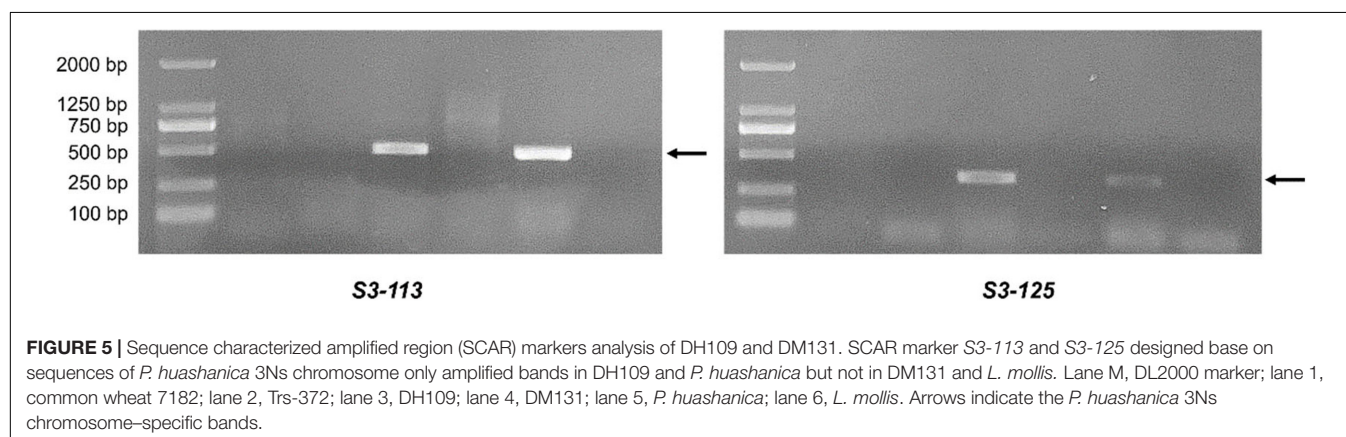
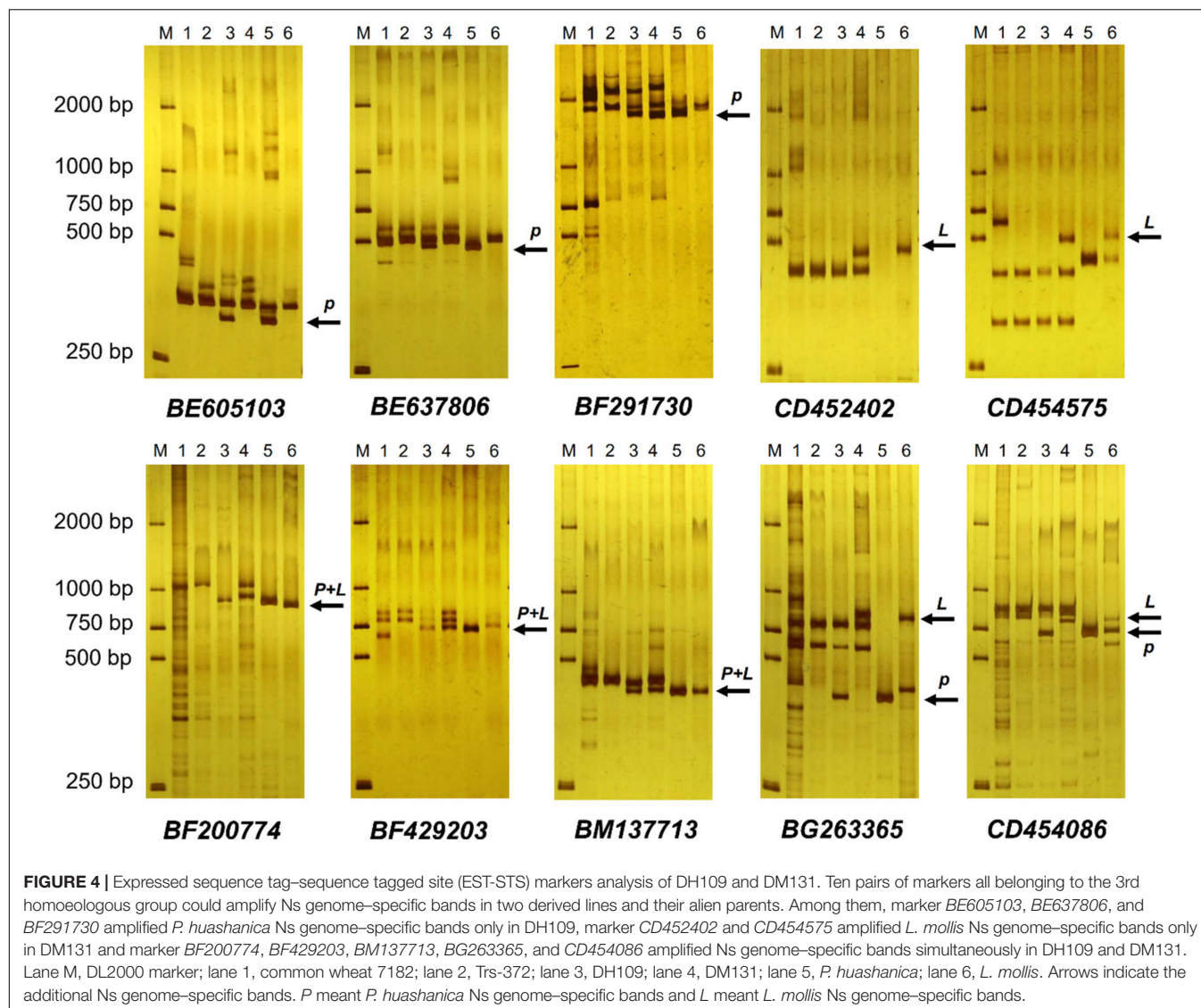
Ns genome-specific bands indicated with arrows in Figure 4 amplified in the two lines and their alien parents by EST-STS markers *BF200774*, *BF429203*, and *BM137713* were sequenced. The results showed that DH109 had identical nucleotide sequences with *P. huashanica*, and DM131 was identical to *L. mollis*. Sequences from 3Ns chromosome between *P. huashanica* and *L. mollis* have high similarity. Compared with wheat 21 pairs of chromosomes, these sequences of specific bands from 3Ns chromosomes only matched with the wheat 3rd homoeologous group chromosomes (Table 2).

FISH and Sequential GISH Analysis of DH109 and DM131

Fluorescence *in situ* hybridization analysis used oligo-primer pSc119.2 and pTa535-1 (Table 1) to distinguish the substituted wheat chromosomes in DH109 and DM131 when compared FISH fluorescent karyotype of the two lines with standard patterns of common wheat Chinese Spring (Annamaria et al., 2003; Tang et al., 2014). In line DH109, a pair of chromosomes with no fluorescent signal replaced wheat 3D chromosomes, which should have red and green signals (Figure 6A, arrows). Sequential GISH analysis was conducted in the slide and indicated that the pair of no-signal chromosomes belonged to *P. huashanica* chromosomes (Figure 6B, arrows). In line DM131, a pair of 3D chromosomes was absent, and another pair of chromosomes with entirely new karyotype appeared in wheat background (Figure 6C, arrows). In sequential GISH experiment, the pair of chromosomes exhibited

TABLE 1 | Molecular markers and FISH oligo-primers used in this study to analyze the chromosomal composition of DH109 and DM131.

Marker	Type	Primer (5'–3')	Tm (°C)	Location
BE605103	STS	F: ACCGACATCACCCATGTCTT, R: CGGCATAGACGGATAGGCT	60	3AL 3BL 3DL
BE637806	STS	F: TCGCAGATCTTCGTTGTTTG, R: GGAATGTGTGGATATTCGG	60	3AS 3BS 3DS
BF291730	STS	F: TTAAGAACCCCAACCCACAGC, R: AGCAGCGCACGGTATTACT	60	3AL 3B 3DL
CD452402	STS	F: ACATACACCCCTCTTGCCGTC, R: GCTTCTTCAAAAGGGCAGTG	60	3AL 3BL 3DL
CD454575	STS	F: AAGGGGTACCCGCATAATTC, R: TCTGAGATACCAGGATGGC	60	3BS 3DS
BF200774	STS	F: AGTTCCTCAGCGTGTGCCTT, R: AATGTGGTGTTCATGGGGAT	60	3AL 3BL 3DL
BF429203	STS	F: CTTGCTAGCCTCCTCACTGG, R: AGATTATGTGCGTGCTGTGC	60	3AL 3BL 3DL
BM137713	STS	F: CTGTCCTTGTAATGGTCCCTG, R: AGGTAAAGCCCGTTCCGT	60	3AL 3BL 3DL
BG263365	STS	F: AACTATCGATGAGATGCGGG, R: GAAGCCTTGAGACCTCCTT	60	3AL 3BL 3DL
CD454086	STS	F: CTCTGTGTGTGGCACTCCAT, R: ATTGCTCGAGATGATGGGTC	60	3AL 3BL 3DL
barc135	SSR	F: ATCGCCATCTCCTCTACCA, R: GCGAACCCATGTGCTAAGT	52	3DS
xcfd64	SSR	F: ACAGTGTTGTTGCCCTTTTC, R: CCCATGTTACAGCTTTGGGT	60	3DL
xcfd223	SSR	F: AAGAGCTACAATGACCAGCAGA, R: GCAGTGTATGTCAGGAGAAGCA	60	3DS
xgwm52	SSR	F: CTATGAGGCGGAGGTTGAAG, R: TGCGGTGCTCTTCCATTT	60	3DL
xgwm314	SSR	F: AGGAGCTCCTCTGTGCCAC, R: TTCGGACTCTCTCCCTG	55	3DS
xgwm456	SSR	F: TCTGAACATTACACAACCCTGA, R: TGCTCTCTCTGAACCTGAAGC	55	3DL
xgwm497	SSR	F: GTAGTGAAGACAAGGGCATT, R: CCGAAAGTTGGGTGATATAC	55	3DS
Xgwm645	SSR	F: TGACCGGAAAAGGGCAGA, R: GCCCTGTCAGGAGTTTAAGT	55	3DL
S3-113	SCAR	F: CGAATTGGATTGGCAGAGGGA, R: ACGATCTCCCTACGAATTGCA	60	5Ns
S3-125	SCAR	F: GGTGACGAGGGTGTGGATG, R: AGTGAACCGCATGGGTCTTT	58	5Ns
pSc119.2	Oligo-probe	6-FAM-CCGTTTTGTGGACTATTACTACCCGCTTTGGGGTCCCAGCTAT	55	–
pTa535-1	Oligo-probe	Tamra-AAAACTTGACGCACGTACGTACAATTGGACAACTCT TCGG AGTAT CAGGGTTTC	55	–



hybridization signals, which meant they were chromosomes from *L. mollis* (Figure 6D, arrows). Combining with the results of molecular markers, wheat 3D chromosomes were replaced by

P. huashanica 3Ns chromosomes in DH109 and by *L. mollis* 3Ns chromosomes in DM131. Therefore, line DH109 was a wheat–*P. huashanica* 3Ns (3D) disomic substitution line,

TABLE 2 | Analysis of the Ns genome-specific bands amplified in DH109 and DM131 by three EST-STS markers.

Adopted markers	Chromosomal source of products	Size of products (bp)	GenBank accession number	Best match with sequences in wheat chr (percentage)	Alignment of the two genomes
BF200774	<i>P. huashanica</i> 3Ns	856	MW114947	3D (87%)	86%
	<i>L. mollis</i> 3Ns	864	MW114948	3B (90%)	
BF429203	<i>P. huashanica</i> 3Ns	674	MW114949	3A (76%)	74%
	<i>L. mollis</i> 3Ns	671	MW114950	3D (83%)	
BM137713	<i>P. huashanica</i> 3Ns	464	MW114951	3D (87%)	92%
	<i>L. mollis</i> 3Ns	456	MW114952	3D (87%)	

Each marker amplified similar-size bands in different genomes.

and line DM131 was a wheat-*L. mollis* 3Ns (3D) disomic substitution line.

Wheat 55K SNP Array Analysis of Two Lines

The wheat 55K SNP arrays were used for comparison of fingerprints. When compared with their parents, DH109 exhibited higher similarity with shared wheat parent line 7,182 than DM131 in terms of percentage of same SNP loci in 21-pair chromosomes, and both derived lines showed low similarity with their respective alien parents (Table 3). There was an obvious commonality that cross points were in 3D chromosomes in which two lines had minimum probeset loci allele with wheat parent but had most of the same allele as their respective alien parents at corresponding position (Figures 7A,B). To make comparisons objective, the valid SNPs were arranged in 3D chromosome basing their physical position (Figures 7C,D). The results showed that these SNPs distributed evenly on the entire 3D chromosome, and both derived lines expressed more same alleles in the same positions as their respective alien parents rather than wheat parent 7,182.

Differences in Diseases Resistance and Agronomic Traits of DH109 and DM131

The response of materials to wheat stripe rust at adult stage was tested in the field and all materials grown under the same condition to ensure the accuracy of results. The ITs of the seven materials were as follows: susceptible control Mingxian 169 (IT = 4), line 7,182 (IT = 3), Trs-372 (IT = 2), DH109 (IT = 1), DM131 (IT = 2), *L. mollis* (IT = 0), and *P. huashanica* (IT = 0) (Figure 8A). This indicated that the introduced alien chromosomes did not make the two lines exhibit great resistance to mixed *Pst* races (CYR32, 33, 34).

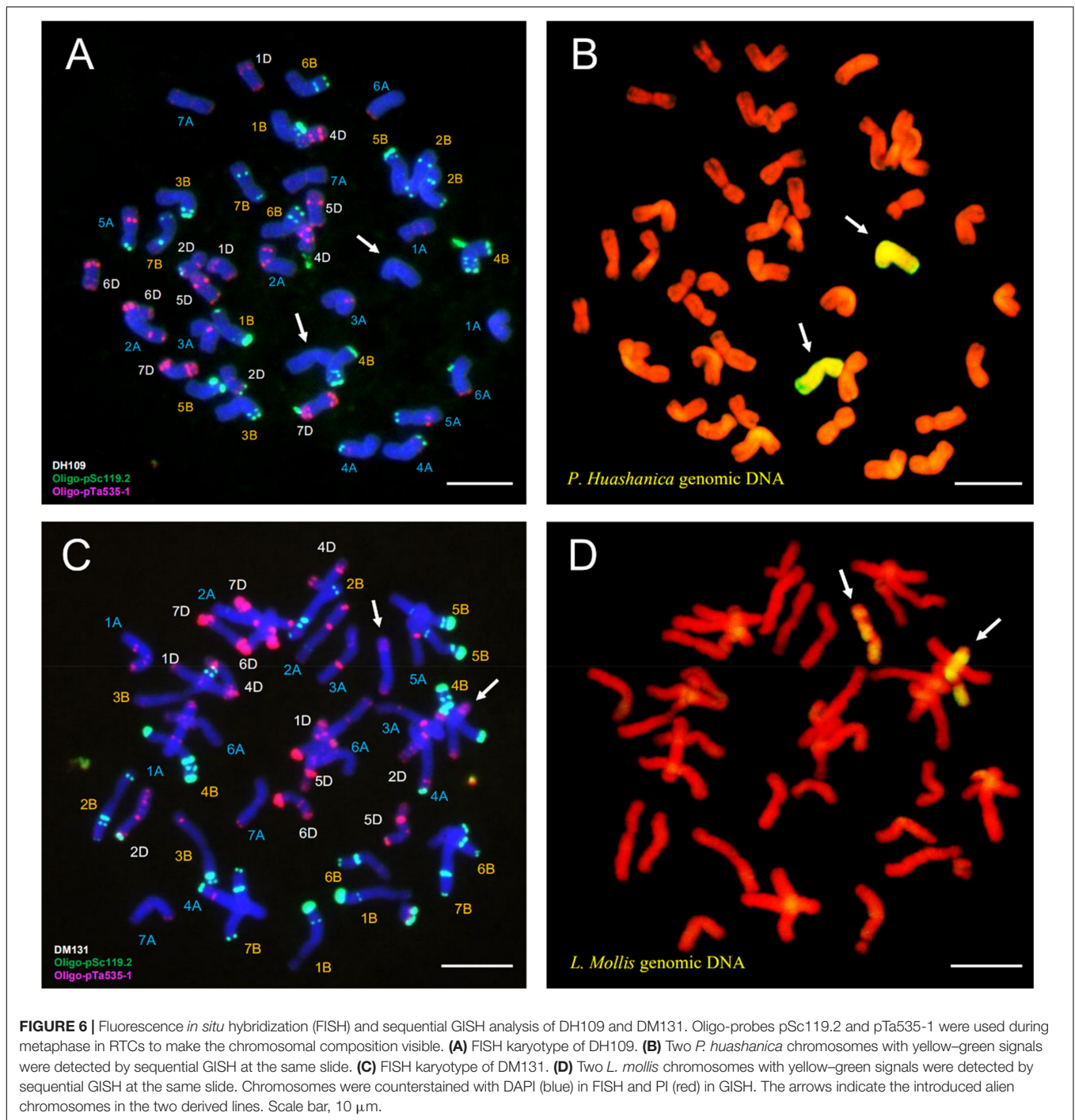
Resistance to powdery mildew was evaluated in growth chamber in the seedling age and infected using *Bst* isolate E09. The ITs of seven materials were as follows: susceptible control Huixianhong (IT = 4), line 7182 (IT = 3+), Trs-372 (IT = 3), DH109 (IT = 0), DM131 (IT = 0), *L. mollis* (IT = 0), and *P. huashanica* (IT = 0) (Figure 8B). It is obvious that DH109 and DM131 were almost immune to inoculated *Bst* isolate, indicating that both lines acquired powdery mildew resistance genes from their alien parent.

The spikelets that kraurotic or covered with mycelium were considered infected after injection. Among them, only DH109 expressed high FHB resistance (RI = 1.57, ISR = 9.86%); susceptible control and other materials all had severe symptoms: CS (RI = 4.87, ISR = 93.87%), 7182 (RI = 4.4, ISR = 70.58%), Trs-372 (RI = 4.64, ISR = 76.74%), and DM131 (RI = 4.37, ISR = 80.25%) (Figure 8C). In response to FHB strain PH1, DH109 and DM131 expressed visible difference that DH109 was superior to DM131.

The morphological traits of the two substitution lines and their wheat parents (7182 and Trs-372) could be seen in Figures 8D–F and Table 4. DH109 and DM131 both exhibited shorter plant compared with their wheat parents 7182 and Trs-372, but DH109 had bigger kernels, and DM131 had longer spikes, more kernels per spike, and tiller number (at $p = 0.05$ and $p = 0.01$). However, grain quality indicators showed that the two lines fell in between common wheat 7182 and durum wheat Trs-372, meaning grain quality of DH109 and DM131 had no significant improvement (Table 5).

DISCUSSION

The most common way to utilize the genome of relative species of wheat is to create wheat-alien-derived lines, including addition lines, substitution lines, translocation lines, and introgression lines, and then cross plus multiple backcross with wheat varieties to obtain new wheat germplasms containing objective traits of alien species (Li et al., 2008). The Ns genome consists of seven homoeologous groups (1–7Ns) that all have been testified useful for wheat breeding improvement because of many beneficial genes: leaf rust resistance genes located in 1Ns, 3Ns, and 7Ns (Du et al., 2012, 2014c; Pang et al., 2014); stripe rust resistance genes located in 2Ns, 3Ns, 4Ns, 5Ns, and 7Ns (Du et al., 2012, 2014a,b,c; Li J. C. et al., 2019; Li et al., 2020a); powdery mildew resistance genes located in 1Ns and 5Ns (Han et al., 2020; Li et al., 2020b); spike characters-related genes located in 4Ns and 6Ns (Du et al., 2013, 2014a); and gluten synthesis-related genes located in 1Ns, 5Ns, and 6Ns (Zhao et al., 2010; Du et al., 2013; Li J. C. et al., 2019). Years of research and numerous evidences, i.e., chromosomal pairing in meiotic stage, molecular markers studies, southern hybridization, and GISH analysis, all show that *Leymus* share the same Ns genome from *Psathyrostachys*, whereas Ns genome in *Leymus* ought to be a mutational version



(Wang et al., 2006; Yen et al., 2009). In the long course of evolution, variation in genetic material is inevitably large, which makes species from different genera appear to be closer than they are within the same genus in the cluster analysis of the *Psathyrostachys*–*Leymus* group according to the results of restriction fragment length polymorphism (Anamthawat-Jonsson and Bodvarsdottir, 2001). In this case, are there big differences in agronomic traits if Ns genome chromosomes from the same homoeology but different genera are separately

introduced to wheat background? Therefore, wheat alien–derived lines with the same homoeologous chromosomes were created for such comparison. In the current study, we developed and identified two novel wheat–alien substitution lines, which one carried a pair of *P. huashanica* 3Ns chromosomes and one carried a pair of *L. mollis* 3Ns chromosomes. Both pairs of the Ns chromosomes are stable and inheritable and caused obvious changes of their wheat receptor parent in agronomic traits and disease resistance.

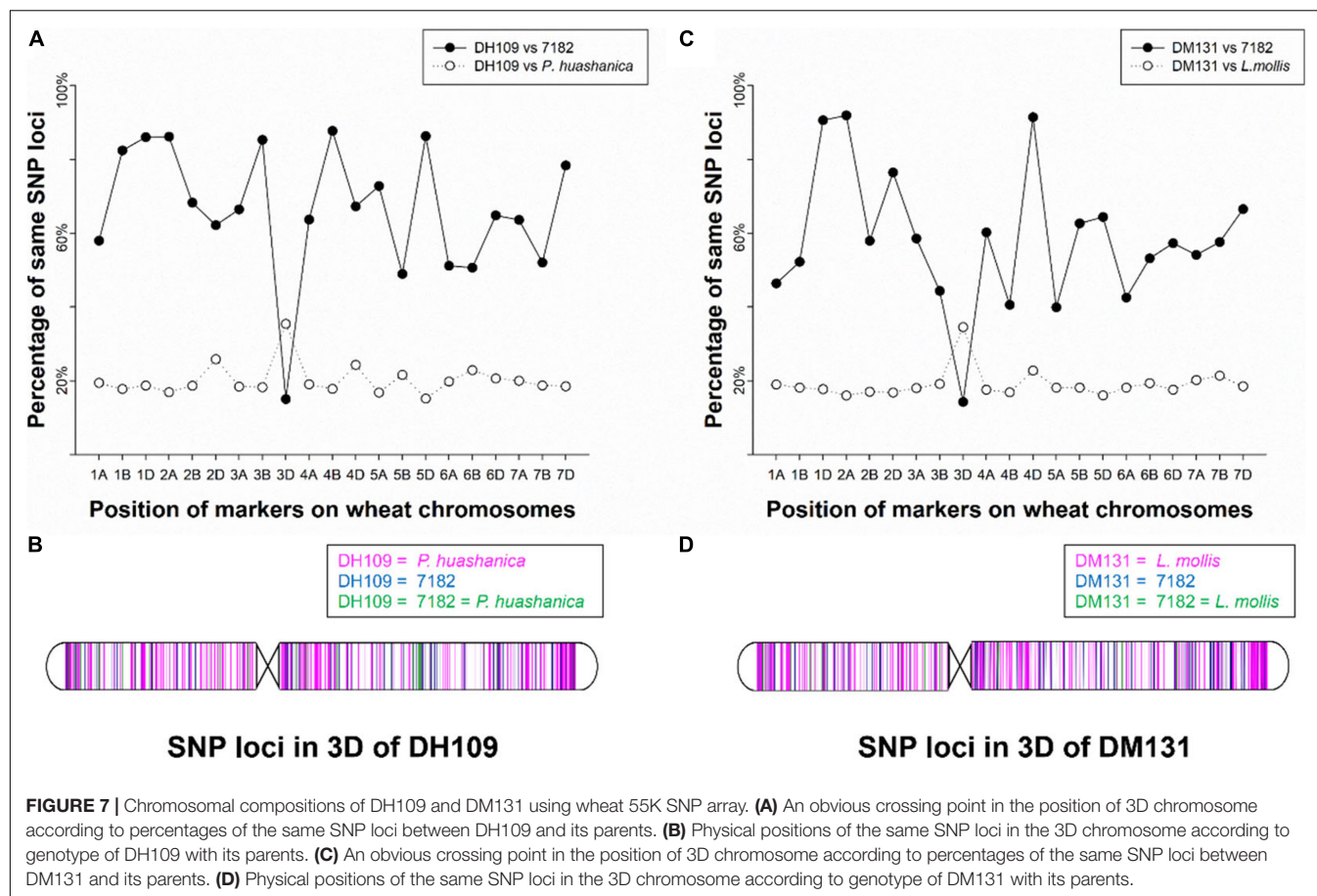
TABLE 3 | Comparison of wheat 55K SNP array data between the two derived lines and their wheat parent 7182.

Chromosome	Number of valid markers	Number of same markers (7182 vs. DH109)	Percentage of same markers (7182 vs. DH109)	Number of same markers (DH109 vs. P. h.)	Percentage of same markers (DH109 vs. P. h.)	Number of same markers (7182 vs. DM131)	Percentage of same markers (7182 vs. DM131)	Number of same markers (DM131 vs. L. m.)	Percentage of same markers (DM131 vs. L. m.)
1A	2,624	1,521	58.0%	509	19.4%	1,217	46.4%	497	18.9%
1B	2,595	2,138	82.4%	460	17.7%	1,355	52.2%	470	18.1%
1D	2,138	1,839	86.0%	399	18.7%	1,937	90.6%	378	17.7%
2A	2,622	2,259	86.2%	442	16.9%	2,410	91.9%	419	16.0%
2B	2,600	1,775	68.3%	484	18.6%	1,506	57.9%	442	17.0%
2D	2,247	1,397	62.2%	580	25.8%	1,718	76.5%	377	16.8%
3A	2,174	1,444	66.4%	400	18.4%	1,273	58.6%	391	18.0%
3B	2,595	2,213	85.3%	473	18.2%	1,150	44.3%	496	19.1%
3D	1,693	254	15.0%	599	35.4%	241	14.2%	584	34.5%
4A	2,592	1,650	63.7%	493	19.0%	1,561	60.2%	455	17.6%
4B	2,556	2,242	87.7%	454	17.8%	1,036	40.5%	430	16.8%
4D	1,420	955	67.3%	345	24.3%	1,298	91.4%	323	22.7%
5A	2,611	1,900	72.8%	438	16.8%	1,041	39.9%	473	18.1%
5B	2,586	1,266	49.0%	558	21.6%	1,620	62.6%	468	18.1%
5D	1,737	1,499	86.3%	263	15.1%	1,118	64.4%	278	16.0%
6A	2,623	1,342	51.2%	518	19.7%	1,115	42.5%	475	18.1%
6B	2,547	1,290	50.6%	582	22.9%	1,354	53.2%	492	19.3%
6D	1,728	1,120	64.8%	356	20.6%	990	57.3%	303	17.5%
7A	2,579	1,640	63.6%	515	20.0%	1,396	54.1%	520	20.2%
7B	2,487	1,294	52.0%	465	18.7%	1,432	57.6%	531	21.4%
7D	2,305	1,806	78.4%	426	18.5%	1,533	66.5%	426	18.5%
A genome	17,825	11,756	66.0%	3,315	18.6%	10,013	56.2%	3,230	18.1%
B genome	17,966	12,218	68.0%	3,476	19.3%	9,453	52.6%	3,329	18.5%
D genome	13,268	8,870	66.9%	2,968	22.4%	8,835	66.6%	2,669	20.1%
Total	49,059	32,844	66.9%	9,759	19.9%	28,301	57.7%	9,228	6.7%

Classical cytogenetics is a necessary way to give an insight into chromosomal composition and transmission of materials. When wheat-alien-derived lines have consistent traits in several successive years, the compositions and behaviors in their RTCs and PMCs need to be observed (Cifuentes and Benavente, 2009). In this study, observations showed that DH109 and DM131 both had 42 chromosomes in somatic cells, and they could pair up to form 21 bivalents in meiotic metaphase I. Subsequently, half of the chromosomes, respectively, moved to cell pole without lagging in anaphase I. GISH technology was first applied to identify alien chromosome(s) in wheat in 1989 and improved by Le et al. (1989) and Mukai and Gill (1991). Since then, the content and behavior of alien chromosome in wheat background could be visualized. GISH analysis suggested that DH109 had 40 wheat chromosomes plus two *P. huashanica* chromosomes, and the two chromosomes were from one homoeology because of their behaviors in pairing, segregation, and transmitting. Same as in DH109, two *L. mollis* chromosomes in DM131 were from one homoeology and stably inherited.

We only knew that a pair of wheat chromosomes was substituted by a pair of alien chromosomes in the two lines through GISH experiments; therefore, genome-specific molecular markers were adopted to determine homoeology of these alien chromosomes and the lost wheat chromosomes.

A wheat-*L. mollis* 2Ns, 3Ns (2D, 3D) double substitution line and a wheat-*P. huashanica* 5Ns (5D) substitution line were identified by Li J. C. et al. (2019) and Zhao et al. (2019) using SSR, EST-STS, and SCAR markers. SSR markers located at a specific position of wheat chromosome could be regarded as tags for the presence of chromosomes (arms) (Röder et al., 1998). EST-STS markers were designed from coding DNA and were generally highly conserved, so they could be used for comparative genomic studies between distant species (Kamaluddin et al., 2017). In this study, we found that SSRs on 3D chromosomes amplified D genome-specific bands in wheat parent 7182, but not in lines DH109 and DM131. EST-STSs on the third homoeologous group amplified Ns genome-specific bands in DH109, DM131, and alien species. SCARs were designed based on sequences of *P. huashanica*-amplified bands only in materials containing *P. huashanica* 3Ns chromosomes. So, it could be preliminarily determined that DH109 was a wheat-*P. huashanica* 3Ns (3D) substitution line, and DM131 was a wheat-*L. mollis* 3Ns (3D) substitution line. It was worth noticing that although EST-STSs and SCARs from third homoeologous group could be used to identify 3Ns chromosomes, the target bands might be different in different species, e.g., *P. huashanica* and *L. mollis*, which showed the same homoeologous chromosomes containing different genetic



materials in different genera. Parallel bands sequencing results indicated that 3rd homoeologous group chromosomes from the three genera had high homoeology, which mainly showed up as dispersedly multiple-bases differences rather than continuous differences in long segments.

Fluorescence *in situ* hybridization analysis was an efficient and reliable way to detect structural rearrangements and replacements of wheat chromosomes because appropriate match of FISH oligo-probers could distinguish 21 pairs of wheat chromosomes according to the standard FISH karyotypes (Huang et al., 2018). The most commonly used FISH probes were oligo-GAA (A and B genome), oligo-pSc119.2 (B genome), oligo-pTa535 (A and D genome), and oligo-pAs1 (A and D genome) (Danilova et al., 2012; Tang et al., 2014). In today's study, matching of pSc119.2 and pTa535 was employed, and the results showed that DH109 and DM131 both lost their 3D chromosomes but possessed a pair of chromosomes whose FISH karyotypes were brand new. Therefore, sequential GISH was conducted to further characterize these exceptional chromosomes in the same slide. With the results of molecular markers, the chromosomes with novel karyotypes in DH109 were *P. huashanica* 3Ns chromosomes and in DM131 were *L. mollis* 3Ns chromosomes. Comparing oligo-probe pSc119.2 and pTa535-1 FISH pattern and ideogram of third homoeologous chromosomes from different genera, big differences could be

seen in **Figure 9**. It was clear that the terminal part of chromosome arms exhibited red and green fluorescent signals in wheat 3D chromosome and red fluorescent signals in *L. mollis* 3Ns chromosomes. However, none of the signals were in *P. huashanica* 3Ns chromosomes. Repetitive sequences have been estimated to be between 16 and 45% in cereal genome, which are helpful in differentiating closely related species, detecting interspecific hybrids and introgressions (Anamthawat-Jónsson and Heslop-Harrison, 1993). Therefore, the results demonstrated that the relationship between *L. mollis* and *Triticum* was closer than *P. huashanica*, and *L. mollis* had distant phylogenetic relationship with *P. huashanica*, which supported the inferences (i.e., donor species of Ns genome to *Leymus* was not *P. huashanica*) of Bodvardsdottir and Anamthawat-Jonsson (2003) and Wang et al. (2006) based on their studies of SSR markers and Southern blot.

Molecular marker-assisted selection has three main categories comprising markers based on molecular hybridization, e.g., restriction fragment length polymorphism and variable number of tandem repeat; markers based on PCR technology, e.g., random amplified polymorphism DNA and SSR; and markers based on high-throughput DNA sequence, e.g., SNP arrays and specific-locus amplified fragment sequencing (SLAF-seq) (He et al., 2003; Varshney et al., 2005; Kamaluddin et al., 2017). Among them, SNP arrays are likely to be the most important

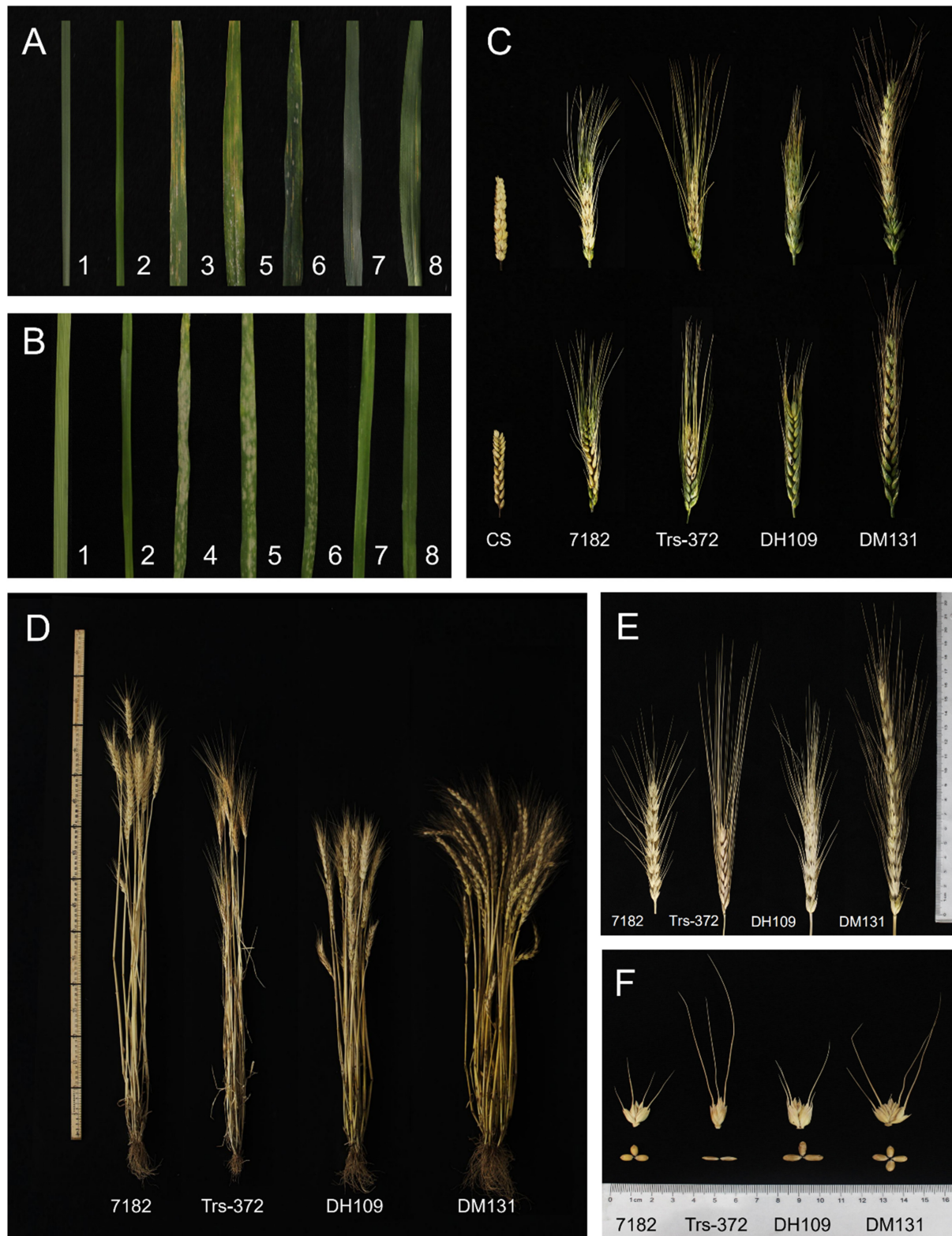


FIGURE 8 | Diseases resistance and agronomic traits of DH109 and DM131, their wheat parents, and controls. **(A)** Symptoms in response to inoculation with a mixture of *Pst* races in the adult stage. **(B)** Symptoms in response to inoculation with a mixture of *Bgt* isolate E09 in the seedling stage. **(C)** Spike symptoms after injection with *Fusarium graminearum*. **(D)** Plants of materials. **(E)** kernels of materials. **(F)** Spikes of materials. The materials in the Figure are 1, *L. mollis*; 2, *P. huashanica*; 3, control Mingxian 169; 4, control Huixianhong; 5, common wheat 7182; 6, durum wheat Trs-372; 7, line DH109; 8, line DM131.

TABLE 4 | Morphological traits of common wheat 7182, durum wheat Trs-372, DH109, and DM131.

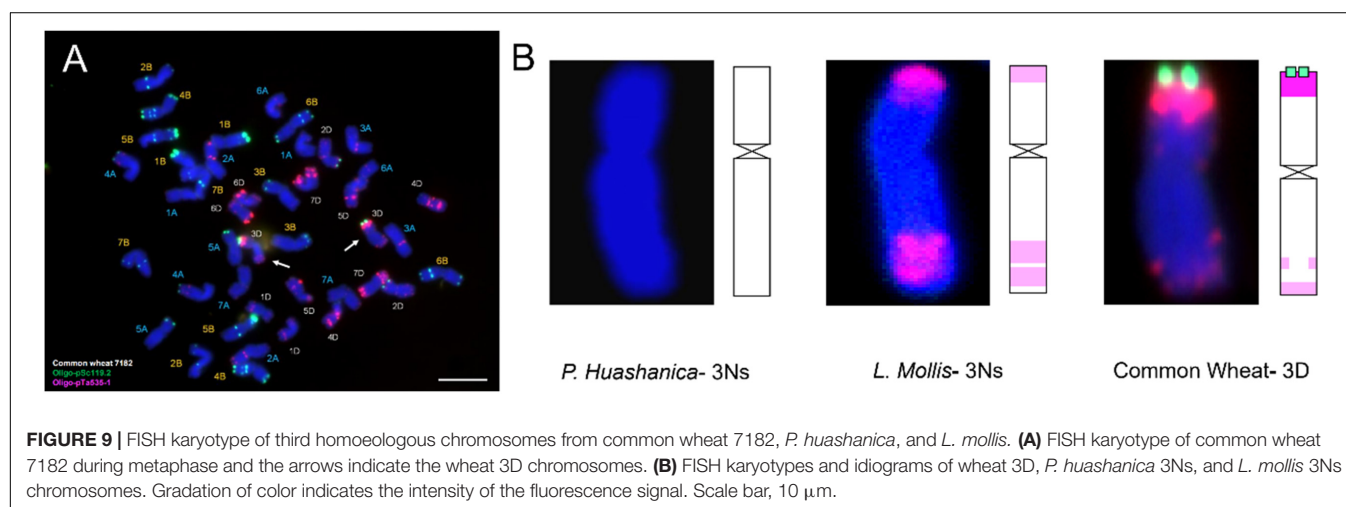
Material	Plant height (cm)	Tiller number	Spike length (cm)	Spikelets per spike	Kernels per spike	Thousand kernel weight (g)
7182	79.9 ± 3.3Aa	12 ± 2BCbc	9.04 ± 0.59Bb	18 ± 2Bb	55 ± 4Bb	37.78 ± 0.82Bb
Trs-372	70.7 ± 2.6Bb	11 ± 2Cc	6.87 ± 0.4Cc	14 ± 3Cc	39 ± 3Cc	32.18 ± 0.94Cc
DH109	62.3 ± 2.1Cd	14 ± 2Bb	8.5 ± 0.62Bb	17 ± 2BCb	55 ± 3Bb	43.66 ± 0.61Aa
DM131	65.7 ± 1.9Cc	28 ± 2Aa	16.46 ± 1.01Aa	24 ± 3Aa	71 ± 3Aa	38.25 ± 0.82Bb

Different uppercase and lowercase letters indicate significant differences at $p < 0.01$ and $p < 0.05$, respectively, between the two substitution lines and their parents.

TABLE 5 | Grain quality results of DH109, DM131, and their wheat parents.

Material	Crude protein content	Gluten protein content	Starch content	Subsidence value	Volume-weight	Dough stability time	Flour field
7182	13.04 ± 0.26Cc	25.33 ± 1.01Cc	56.49 ± 0.765Bb	24.41 ± 1.13Cc	788.5 ± 6Aa	2.2 ± 0.7Bc	72 ± 2ABb
Trs-372	16.67 ± 0.45Aa	34.44 ± 0.99Aa	63.79 ± 1.42Aa	41.19 ± 1.07Aa	786 ± 2Aa	6.8 ± 1.6Aa	76 ± 2Aa
DH109	14.47 ± 0.32Bb	34.25 ± 0.57Aa	62.73 ± 0.98Aa	34.85 ± 1.53Bb	754.5 ± 4Bb	4.9 ± 0.3Ab	66 ± 1Cd
DM131	14.3 ± 0.54BCb	30.47 ± 0.61Bb	58.36 ± 0.91Bb	24.695 ± 1.77Cc	757 ± 2Bb	1.4 ± 0.7Bc	68.5 ± 1BCc

Different uppercase and lowercase letters indicate significant differences at $p < 0.01$ and $p < 0.05$, respectively, between the two substitution lines and their parents.



tool in gene mapping and study the relationship of species (Zhou et al., 2018; Zhao et al., 2019). Wheat SNP arrays were first applied in identification of wheat-alien-derived line by Li J. C. et al. (2019) and Li et al. (2020b) and successfully verified two derived lines. In the present study, we compared the genotype of DH109 with its parents and DM131 with its parents in each locus of wheat 55K SNPs that spread on 21 chromosomes, respectively. It was consistent with results from molecular markers and FISH analysis, which showed DH109 was a wheat-*P. huashanica* 3Ns (3D) substitution line and DM131 was a wheat-*L. mollis* 3Ns (3D) substitution line. Compared with the results of previous articles of 15K SNP array in this section, it can be obviously found that low-density SNP array got higher resolution in substitution-occurred or translocation-occurred homoeologous group. Therefore, in the identification of exogenous substances, it might be more efficient and cheaper to use 15K or 35K SNP arrays rather than 90K or 660K SNP arrays.

The introduction and replacement of alien chromosome(s) or segment(s), except for B-chromosomes, usually cause changes

in the traits of recipient plant (Camacho et al., 2000; Jones et al., 2008). In wheat, these changes might be obvious in morphologic traits, such as plant height (Wang S. W. et al., 2019) and kernel size (Zhang et al., 2016); they also might be invisible in resistance or grain quality (He et al., 2017; Li X.Y. et al., 2019). Unfortunately, not all alien chromosomes were beneficial to wheat because they might result in worse agronomic traits, e.g., small spike and less tiller (Wang J. et al., 2019), and decreased processing quality (Liu et al., 2004). Therefore, return to breeding requirement, the most important criterion to access value of one wheat-alien-derived line, was its agronomic trait. In this study, two substitution lines DH109 and DM131 both expressed high resistance to powdery mildew in their seeding age. Moreover, DH109 also had high FHB resistance and bigger kernels, and DM131 had longer spike and more tiller number, which were outstanding agronomic traits for wheat improvement. Although the two derived lines both possessed a pair of alien chromosomes that belonged to the 3rd homoeology and named 3Ns, they had obviously different agronomic traits because DH109 had

P. huashanica 3Ns chromosomes, and DM131 had *L. mollis* 3Ns chromosomes. The chromosomal recombination and crossing with durum wheat Trs-372 might cause individual difference even in the same generation; for example, DH109 was more like common wheat 7182 in plant type, and DM131 was more like durum wheat Trs-372 in grain quality. However, enhanced/increased disease resistance and some excellent traits are most likely caused by introduction of alien chromosomes. Relatively large differences in agronomic traits between DH109 and DM131 supported that there were many different genes in Ns genome between *P. huashanica* and *L. mollis*.

CONCLUSION

In this study, a novel wheat–*P. huashanica* disomic substitution line named DH109 and a novel wheat–*L. mollis* disomic substitution line named DM131 were identified by using molecular and cytogenetic methods. Although both two lines were developed because of the substitution of exogenetic 3Ns chromosomes and wheat 3D chromosomes, they were obviously different in bands of molecular markers, FISH karyotype and agronomic traits. Thus, Ns genome from *P. huashanica* and *L. mollis* had big differences. Furthermore, after multiple generation advancement, the two lines have been stable in morphology and genetics. Line DH109 expressed superior resistance to powdery mildew and FHB, and line DM131 had powdery mildew resistance, longer spike, and more tiller number. Therefore, the two lines could have different preferences toward wheat breeding and Ns genome research.

REFERENCES

- An, D. G., Zheng, Q., Zhou, Y. L., Ma, P. T., Lv, Z. L., Li, L. H., et al. (2013). Molecular cytogenetic characterization of a new wheat–rye 4R chromosome translocation line resistant to powdery mildew. *Chromosome Res.* 21, 419–432. doi: 10.1007/s10577-013-9366-8
- Anamthawat-Jonsson, K., and Bodvardsdottir, S. (2001). Genomic and genetic relationships among species of *Leymus* (Poaceae: Triticeae) inferred from 18S–26S ribosomal genes. *Am. J. Bot.* 88, 553–559. doi: 10.2307/2657053
- Anamthawat-Jonsson, K., and Heslop-Harrison, J. S. (1993). Isolation and characterization of genome-specific DNA sequences in *Triticeae* species. *Mol. Gen. Genet.* 240, 151–158. doi: 10.1007/bf00277052
- Annamaria, S., Linc, G., Molnar-Lang, M., and Graner, A. (2003). Fluorescence *in situ* hybridization polymorphism using two repetitive DNA clones in different cultivars of wheat. *Plant Breed.* 122, 396–400. doi: 10.1046/j.1439-0523.2003.00891.x
- Baden, C. (1991). A taxonomic revision of *Psathyrostachys* (Poaceae). *Nordic J. Bot.* 11, 3–26. doi: 10.1111/j.1756-1051.1991.tb01790.x
- Bai, G. H. (1996). Variation in *Fusarium graminearum* and cultivar resistance to wheat scab. *Plant Dis.* 80:975. doi: 10.1094/PD-80-0975
- Bai, G. H., Kolb, F. L., Shaner, G., and Domier, L. L. (1999). Amplified fragment length polymorphism markers linked to a major quantitative trait locus controlling scab resistance in wheat. *Phytopathology* 89, 343–348. doi: 10.1094/phyto.1999.89.4.343
- Bai, G. H., Shaner, G., and Ohm, H. (2000). Inheritance of resistance to *Fusarium graminearum* in wheat. *Theor. Appl. Genet.* 100, 1–8. doi: 10.1007/PL00002902

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

JCL and JJJL conducted the experiments. LZ and XNC analyzed the data. JW and QY contributed new reagents and analytical tools. ZY contributed new methods. JZ and XHC conceived and designed the research. JCL wrote the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (31771785) and Key Projects in Shaanxi Provincial Agricultural Field (2018ZDXM-NY-006).

ACKNOWLEDGMENTS

We would like to thank Yang Liu for her language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.644896/full#supplementary-material>

- Bodvardsdottir, S., and Anamthawat-Jonsson, K. (2003). Isolation, characterization, and analysis of *Leymus*-specific DNA sequences. *Genome* 46, 673–682. doi: 10.1139/g03-029
- Camacho, J. P. M., Sharbel, T. F., and Beukeboom, L. W. (2000). B-chromosome evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 163–178. doi: 10.1098/rstb.2000.0556
- Chantret, N., Salse, J., Sabot, F., Rahman, S., Bellec, A., Laubin, B., et al. (2005). Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell* 17, 1033–1045. doi: 10.1105/tpc.104.029181
- Chen, L. G., Jun, W. U., Zhao, J. X., Liu, S. H., Yang, Q. H., Wan-Li, D. U., et al. (2010). Screening, cloning and southern blotting analysis of specific repetitive DNA sequences of *Psathyrostachys huashanica*. *J. Triticeae Crops* 30, 23–28. doi: 10.1080/00949651003724790
- Chen, P. D., Qi, L., Zhou, B., Zhang, S., and Liu, D. J. (1995). Development and molecular cytogenetic analysis of wheat–*Haynaldia villosa* 6VS/6AL translocation lines specifying resistance to powdery mildew. *Theor. Appl. Genet.* 91, 1125–1128. doi: 10.1007/BF00223930
- Chen, S. Y. (1991). The hybridization between *Triticum aestivum* and *Psathyrostachys huashanica*. *Acta Genet. Sin.* 18, 508–512.
- Ciferri, R. (1955). The first interspecific wheat hybrids. *J. Heredity* 46, 81–83. doi: 10.1093/oxfordjournals.jhered.a106529
- Cifuentes, M., and Benavente, E. (2009). Wheat–alien metaphase I pairing of individual wheat genomes and D genome chromosomes in interspecific hybrids between *Triticum aestivum* L. and *Aegilops geniculata* Roth. *Theor. Appl. Genet.* 119, 805–813. doi: 10.1007/s00122-009-1090-6

- Danilova, T. V., Friebe, B., and Gill, B. S. (2012). Single-copy gene fluorescence in situ hybridization and genome analysis: *Acc-2* loci mark evolutionary chromosomal rearrangements in wheat. *Chromosoma* 121, 597–611. doi: 10.1007/s00412-012-0384-7
- Dewey, D. R. (1984). "The genomic system of classification as a guide to intergeneric hybridization with the perennial *Triticeae*," in *Gene Manipulation in Plant Improvement: 16th Stadler Genetics Symposium*, ed. J. P. Gustafson (Boston, MA: Springer US), 209–279. doi: 10.1007/978-1-4613-2429-4_9
- Du, W. L., Wang, J., Lu, M., Sun, S. G., Chen, X. H., Zhao, J. X., et al. (2014a). Characterization of a wheat-*Psathyrostachys huashanica* Keng 4Ns disomic addition line for enhanced tiller numbers and stripe rust resistance. *Planta* 239, 97–105. doi: 10.1007/s00425-013-1957-2
- Du, W. L., Wang, J., Pang, Y. H., Li, Y. L., Chen, X. H., Zhao, J., et al. (2013). Isolation and characterization of a *Psathyrostachys huashanica* Keng 6Ns chromosome addition in common wheat. *PLoS One* 8:e53921. doi: 10.1371/journal.pone.0053921
- Du, W. L., Wang, J., Pang, Y. H., Wang, L. M., Wu, J., Zhao, J. X., et al. (2014b). Isolation and characterization of a wheat-*Psathyrostachys huashanica* 'Keng' 3Ns disomic addition line with resistance to stripe rust. *Genome* 57, 37–44. doi: 10.1139/gen-2013-0199
- Du, W. L., Wang, J., Pang, Y. H., Wu, J., Zhao, J. X., Liu, S. H., et al. (2014c). Development and application of PCR markers specific to the 1Ns chromosome of *Psathyrostachys huashanica* Keng with leaf rust resistance. *Euphytica* 200, 207–220. doi: 10.1007/s10681-014-1145-x
- Du, W. L., Wang, J., Wang, L. M., Zhang, J., Chen, X. H., Zhao, J. X., et al. (2012). Development and characterization of a *Psathyrostachys huashanica* Keng 7Ns chromosome addition line with leaf rust resistance. *PLoS One* 8:e70879. doi: 10.1371/journal.pone.0070879
- Fu, J., Chen, S. Y., and Zhang, A. J. (1993). Studies of the formation and cytogenetics of octoploid *Triticum*. *Acta Genet. Sin.* 20, 317–323.
- Han, H. M., Bai, L., Su, J. J., Zhang, J. P., Song, L. Q., Gao, A. N., et al. (2014). Genetic rearrangements of six wheat-*Agropyron cristatum* 6P addition lines revealed by molecular markers. *PLoS One* 9:e91066. doi: 10.1371/journal.pone.0091066
- Han, J., Liu, Y. X., Hou, C. C., Li, J. C., Wang, J. L., Zhang, Q. Y., et al. (2020). A 1Ns disomic addition from *Psathyrostachys Huashanica* Keng confers resistance to powdery mildew in wheat. *Agronomy(Basel)* 10:312. doi: 10.3390/agronomy10020312
- He, F., Bao, Y. G., Qi, X. L., Ma, Y. X., Li, X. F., and Wang, H. G. (2017). Molecular cytogenetic identification of a wheat-*Thinopyrum ponticum* translocation line resistant to powdery mildew. *J. Genet.* 96, 165–169. doi: 10.1007/s12041-017-0754-2
- He, G. H., Meng, R. H., Newman, M., Gao, G. Q., Pittman, R. N., and Prakash, C. S. (2003). Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol.* 3:3. doi: 10.1186/1471-2229-3-3
- Huang, X. Y., Zhu, M. Q., Zhuang, L. F., Zhang, S. Y., Wang, J. J., Chen, X. J., et al. (2018). Structural chromosome rearrangements and polymorphisms identified in Chinese wheat cultivars by high-resolution multiplex oligonucleotide FISH. *Theor. Appl. Genet.* 131, 1967–1986. doi: 10.1007/s00122-018-3126-2
- Jiang, B., Liu, T. G., Li, H. H., Han, H. M., Li, L. H., Zhang, J. P., et al. (2018). Physical mapping of a novel locus conferring leaf rust resistance on the long arm of *Agropyron cristatum* chromosome 2P. *Front. Plant Sci.* 9:817. doi: 10.3389/fpls.2018.00817
- Jing, J. X., Fu, J., Yuan, H. X., Wang, M. N., Shang, H. S., and Li, Z. Q. (1999). A preliminary study on heredity of the resistance to stripe rust in three wild relatives of wheat. *Acta Phytopathol. Sin.* 29, 147–150.
- Jones, R. N., Viegas, W., and Houben, A. (2008). A century of B chromosomes in plants: so what? *Ann. Bot.* 101, 767–775. doi: 10.1093/aob/mcm167
- Kamaluddin, A. A., Khan, M. A., Kiran, U., Ali, A., Abdin, M. Z., Zargar, M. Y., et al. (2017). "Molecular markers and marker-assisted selection in crop plants," in *Plant Biotechnology: Principles and Applications*, eds M. Z. Abdin, U. Kiran, and A. A. Kamaluddin (Singapore: Springer Singapore), 295–328.
- Kishii, M., Wang, R. R. C., and Tsujimoto, H. (2003). Characteristics and behaviour of the chromosomes of *Leymus mollis* and *L. racemosus* (*Triticeae*, *Poaceae*) during mitosis and meiosis. *Chromosome Res.* 11, 741–748. doi: 10.1023/B:CHRO.000005774.00726.71
- Le, H. T., Armstrong, K. C., and Miki, B. (1989). Detection of rye DNA in wheat-rye hybrids and wheat translocation stocks using total genomic DNA as a probe. *Plant Mol. Biol. Rep.* 7, 150–158. doi: 10.1007/BF02669631
- Li, G. R., Lang, T., Dai, G., Li, D. H., Li, C. H., Song, X. J., et al. (2015). Precise identification of two wheat-*Thinopyrum intermedium* substitutions reveals the compensation and rearrangement between wheat and *Thinopyrum* chromosomes. *Mol. Breed.* 35:1. doi: 10.1007/s11032-015-0202-z
- Li, J. C., Liu, Y., Cheng, X. N., Yao, X. N., Yang, Z. J., Wu, J., et al. (2020a). Molecular characteristics and inheritance of a chromosome segment from *Psathyrostachys huashanica* Keng in a wheat background. *Genet. Resour. Crop Evol.* 67, 1245–1257. doi: 10.1007/s10722-020-00908-5
- Li, J. C., Yao, X. N., Yang, Z. J., Cheng, X. N., Yuan, F. P., Liu, Y., et al. (2019). Molecular cytogenetic characterization of a novel wheat-*Psathyrostachys huashanica* Keng 5Ns (5D) disomic substitution line with stripe rust resistance. *Mol. Breed.* 39:109. doi: 10.1007/s11032-019-1014-3
- Li, J. C., Zhao, L., Cheng, X. N., Bai, G. H., Li, M., Wu, J., et al. (2020b). Molecular cytogenetic characterization of a novel wheat-*Psathyrostachys huashanica* Keng T3DS-5NsL5NsS and T5DL-3DS3DL dual translocation line with powdery mildew resistance. *BMC Plant Biol.* 20:163. doi: 10.1186/s12870-020-02366-8
- Li, X. Y., Li, Y., Karim, H., Li, Y., Zhong, X. J., Tang, H. P., et al. (2019). The production of wheat-*Aegilops sharonensis* 15th chromosome substitution lines harboring alien novel high-molecular-weight glutenin subunits. *Genome* 63, 155–167. doi: 10.1139/gen-2019-0106
- Li, Z. S., Li, B., and Tong, Y. P. (2008). The contribution of distant hybridization with decaploid *Agropyron elongatum* to wheat improvement in China. *J. Genet. Genomics* 35, 451–456. doi: 10.1016/s1673-8527(08)60062-4
- Liu, J. J., He, Z., Peña, R., and Zhao, Z. D. (2004). Effect of 1BL/1RS translocation on grain quality and noodle quality in bread wheat. *Acta Agron. Sin.* 30, 149–153. doi: 10.1300/J064v24n01_09
- Löve, Á. (1984). Conspectus of the *Triticeae*. *Feddes Repert.* 95, 425–521. doi: 10.1002/fedr.4910950702
- Ma, H., Singh, R. P., and Mujeeb-Kazi, A. (1995). Suppression/expression of resistance to stripe rust in synthetic hexaploid wheat (*Triticum turgidum* × *T. tauschii*). *Euphytica* 83, 87–93. doi: 10.1007/bf01678034
- Mujeeb-Kazi, A., and Rodriguez, R. (1981). An intergeneric hybrid of *Triticum aestivum* L. × *Elymus giganteus*. *Heredity* 72, 253–256. doi: 10.1093/oxfordjournals.jhered.a109490
- Mukai, Y., and Gill, B. S. (1991). Detection of barley chromatin added to wheat by genomic *in situ* hybridization. *Genome* 34, 448–452. doi: 10.1139/g91-067
- Ørgaard, M., and Heslop-Harrison, J. S. (1994). Relationships between species of *Leymus*, *Psathyrostachys*, and *Hordeum* (*Poaceae*, *Triticeae*) inferred from Southern hybridization of genomic and cloned DNA probes. *Plant Syst. Evol.* 189, 217–231. doi: 10.1007/BF00939728
- Pang, Y. H., Chen, X. H., Zhao, J. X., Du, W. L., Cheng, X. N., Wu, J., et al. (2014). Molecular cytogenetic characterization of a wheat-*Leymus mollis* 3D(3Ns) substitution line with resistance to leaf rust. *J. Genet. Genomics* 41, 205–214. doi: 10.1016/j.jgg.2013.11.008
- Pauk, J. (ed). (2016). Alien introgression in wheat cytogenetics, molecular biology, and genomics. *Cereal Res. Commun.* 44:535. doi: 10.1556/0806.44.2016.041
- Pestsova, E., Ganai, M. W., and Röder, M. S. (2000). Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. *Genome* 43, 689–697. doi: 10.1139/g00-042
- Rabinovich, S. V. (1998). Importance of wheat-rye translocations for breeding modern cultivars of *Triticum aestivum* L. *Euphytica* 100, 323–340. doi: 10.1023/A:1018361819215
- Röder, M., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M., Leroy, P., et al. (1998). A microsatellite map of wheat. *Genetics* 149, 2007–2023. doi: 10.1016/B0-12-227620-5/00113-0
- Sheng, B. Q. (1988). Grades of resistance to powdery mildew classified by different phenotypes of response in the seedling stage of wheat. *Plant Prot.* 1:49.
- Song, S., Tao, Y., Zhang, H., and Wu, Y. (2013). *Psathyrostachys huashanica*, a potential resource for resistance to Barley yellow dwarf virus-GAV. *Eur. J. Plant Pathol.* 137, 217–221. doi: 10.1007/s10658-013-0239-y
- Sourdille, P., Singh, S., Cadalen, T., Brown-Guedira, G. L., Gay, G., Qi, L., et al. (2004). Microsatellite-based deletion bin system for the establishment of genetic-physical map relationships in wheat (*Triticum aestivum* L.). *Funct. Integr. Genomics* 4, 12–25. doi: 10.1007/s10142-004-0106-1

- Su, J. N., Guo, J., Wang, C. J., Jing, F., Zhao, J. X., Yang, Q. H., et al. (2015). Specific SCAR markers on chromosome 3Ns of *Psathyrostachys huashanica* Keng. *J. Triticeae Crops* 35, 1–6. doi: 10.7606/j.issn.1009-1041.2015.01.01
- Sun, G. L., Yen, C., and Yang, J. L. (1995). Morphology and cytology of intergeneric hybrids involving *Leymus multicaulis* (Poaceae). *Plant Syst. Evol.* 194, 83–91. doi: 10.1007/BF00983218
- Tang, Z. X., Yang, Z. J., and Fu, S. L. (2014). Oligonucleotides replacing the roles of repetitive sequences pAs1, pSc119.2, pTa-535, pTa71, CCS1, and pAWRC.1 for FISH analysis. *J. Appl. Genet.* 55, 313–318. doi: 10.1007/s13353-014-0215-z
- Tsitsin, N. V. (1965). Remote hybridisation as a method of creating new species and varieties of plants. *Euphytica* 14, 326–330. doi: 10.1007/BF00149519
- Varshney, R. K., Graner, A., and Sorrells, M. E. (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23, 48–55. doi: 10.1016/j.tibtech.2004.11.005
- Wang, J., Liu, C., Guo, X. R., Wang, K., Du, L. P., Lin, Z. S., et al. (2019). Development and genetic analysis of wheat double substitution lines carrying *Hordeum vulgare* 2H and *Thinopyrum intermedium* 2Ai#2 chromosomes. *Crop J.* 7, 163–175. doi: 10.1016/j.cj.2018.11.003
- Wang, J., Wang, L. M., Du, W. L., Chen, L. G., Liu, S. H., Wu, J., et al. (2014). Development of 5Ns chromosome-specific SCAR markers for utilization in future wheat breeding programs. *Russ. J. Genet.* 50, 606–612. doi: 10.1134/s1022795414060131
- Wang, M., and Shang, H. (2000). Evaluation of resistance in *Psathyrostachys huashanica* to wheat take-all fungus. *Acta Univ. Agric. Boreali-Occidentalis* 28, 69–71.
- Wang, R. R. C., and Hsiao, C. (1984). Morphology and cytology of interspecific hybrids of *Leymus mollis*. *J. Heredity* 75, 488–492. doi: 10.1093/oxfordjournals.jhered.a109992
- Wang, R. R. C., and Jensen, K. B. (1994). Absence of the J genome in *Leymus* (Poaceae: Triticeae): Evidence from DNA hybridization and meiotic pairing. *Genome* 37, 231–235. doi: 10.1139/g94-032
- Wang, R. R. C., Bothmer, R. V., Dvořák, J., Fedak, G., Linde-Laursen, I., and Muramatsu, M. (1994). “Genome symbols in the Triticeae (Poaceae),” in *Proceeding of the 2nd International Triticeae Symposium*, (Logan, UT), 29–34.
- Wang, R. R. C., Zhang, J. Y., Lee, B. S., Jensen, K. B., Kishii, M., and Tsujimoto, H. (2006). Variations in abundance of 2 repetitive sequences in *Leymus* and *Psathyrostachys* species. *Genome* 49, 511–519. doi: 10.1139/G05-126
- Wang, S. W., Wang, C. Y., Wang, Y. Z., Wang, Y. J., Chen, C. H., and Ji, W. Q. (2019). Molecular cytogenetic identification of two wheat-*Thinopyrum ponticum* substitution lines conferring stripe rust resistance. *Mol. Breed.* 39:143. doi: 10.1007/s11032-019-1053-9
- Yang, Z. P., Yang, X. Y., and Huang, D. C. (1998). Studies on somaclonal variants for resistance to scab in bread wheat (*Triticum aestivum* L.) through in vitro selection for tolerance to deoxynivalenol. *Euphytica* 101, 213–219. doi: 10.1023/A:1018354606939
- Yen, C., Yang, J.-L., and Baum, B. (2009). Synopsis of *Leymus* Hochst. (Triticeae: Poaceae). *J. Syst. Evol.* 47, 67–86. doi: 10.1111/j.1759-6831.2009.00004.x
- Zhang, H. B., and Dvořák, J. (1991). The genome origin of tetraploid species of *Leymus* (Poaceae: Triticeae) inferred from variation in repeated nucleotide sequences. *Am. J. Bot.* 78, 871–884. doi: 10.2307/2445166
- Zhang, J., Zhang, J. P., Liu, W. H., Wu, X. Y., Yang, X. M., Li, X. Q., et al. (2016). An intercalary translocation from *Agropyron cristatum* 6P chromosome into common wheat confers enhanced kernel number per spike. *Planta* 244, 853–864. doi: 10.1007/s00425-016-2550-2
- Zhang, Q. P., Li, Q., Wang, X., Wang, H. Y., Lang, S. P., Wang, Y. N., et al. (2005). Development and characterization of a *Triticum aestivum*-*Haynaldia villosa* translocation line T4VS-4DL conferring resistance to wheat spindle streak mosaic virus. *Euphytica* 145, 317–320. doi: 10.1007/s10681-005-1743-8
- Zhao, J. X., Du, W. L., Wu, J., Cheng, X. N., Gao, Y., Pang, Y. H., et al. (2013). Development and identification of a wheat-*Leymus mollis* multiple alien substitution line. *Euphytica* 190, 45–52. doi: 10.1007/s10681-012-0772-3
- Zhao, J. X., Ji, W. Q., Wu, J., Chen, X. H., Cheng, X. N., Wang, J. W., et al. (2010). Development and identification of a wheat-*Psathyrostachys huashanica* addition line carrying HMW-GS, LMW-GS and gliadin genes. *Genet. Resour. Crop Evol.* 57, 387–394. doi: 10.1007/s10722-009-9477-4
- Zhao, J. X., Liu, Y., Cheng, X. N., Pang, Y. H., Li, J. C., Su, Z. Q., et al. (2019). Development and identification of a dwarf wheat-*Leymus mollis* double substitution line with resistance to yellow rust and Fusarium head blight. *Crop J.* 7, 516–526. doi: 10.1016/j.cj.2018.11.012
- Zhou, S. H., Zhang, J. P., Che, Y. H., Liu, W. H., Lu, Y. Q., Yang, X. M., et al. (2018). Construction of *Agropyron Gaertn.* genetic linkage maps using a wheat 660K SNP array reveals a homoeologous relationship with the wheat genome. *Plant Biotechnol. J.* 16, 818–827. doi: 10.1111/pbi.12831
- Zhu, C., Wang, Y. Z., Chen, C. H., Wang, C. Y., Zhang, A. C., Peng, N. N., et al. (2017). Molecular cytogenetic identification of wheat-*Thinopyrum ponticum* substitution line with stripe rust resistance. *Genome* 60, 860–867. doi: 10.1139/gen-2017-0099

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Li, Li, Cheng, Zhao, Yang, Wu, Yang, Chen and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genomic Prediction of Grain Yield in a Barley MAGIC Population Modeling Genotype per Environment Interaction

Damiano Puglisi^{1†}, Stefano Delbono^{2†}, Andrea Visioni³, Hakan Ozkan⁴, İbrahim Kara⁵, Ana M. Casas⁶, Ernesto Igartua⁶, Giampiero Valè⁷, Angela Roberta Lo Piero¹, Luigi Cattivelli², Alessandro Tondelli² and Agostino Fricano^{2*}

¹ Dipartimento di Agricoltura, Alimentazione e Ambiente (Di3A), Università di Catania, Catania, Italy, ² Council for Agricultural Research and Economics—Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy, ³ Biodiversity and Crop Improvement Program, International Center for Agricultural Research in the Dry Areas, Avenue Hafiane Cherkaoui, Rabat, Morocco, ⁴ Department of Field Crops, Faculty of Agriculture, University of Cukurova, Adana, Turkey, ⁵ Bahri Dagdas International Agricultural Research Institute, Konya, Turkey, ⁶ Aula Dei Experimental Station (EEAD-CSIC), Spanish Research Council, Zaragoza, Spain, ⁷ DiSIT, Dipartimento di Scienze e Innovazione Tecnologica, Università del Piemonte Orientale, Vercelli, Italy

OPEN ACCESS

Edited by:

Sukhjiwan Kaur,
Agriculture Victoria, Australia

Reviewed by:

Ezio Portis,
University of Turin, Italy
Sivakumar Sukumaran,
International Maize and Wheat
Improvement Center (Mexico), Mexico

*Correspondence:

Agostino Fricano
agostino.fricano@crea.gov.it

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 04 February 2021

Accepted: 26 April 2021

Published: 24 May 2021

Citation:

Puglisi D, Delbono S, Visioni A,
Ozkan H, Kara İ, Casas AM, Igartua E,
Valè G, Piero AR, Cattivelli L,
Tondelli A and Fricano A (2021)
Genomic Prediction of Grain Yield in a
Barley MAGIC Population Modeling
Genotype per Environment
Interaction.
Front. Plant Sci. 12:664148.
doi: 10.3389/fpls.2021.664148

Multi-parent Advanced Generation Inter-crosses (MAGIC) lines have mosaic genomes that are generated shuffling the genetic material of the founder parents following pre-defined crossing schemes. In cereal crops, these experimental populations have been extensively used to investigate the genetic bases of several traits and dissect the genetic bases of epistasis. In plants, genomic prediction models are usually fitted using either diverse panels of mostly unrelated accessions or individuals of biparental families and several empirical analyses have been conducted to evaluate the predictive ability of models fitted to these populations using different traits. In this paper, we constructed, genotyped and evaluated a barley MAGIC population of 352 individuals developed with a diverse set of eight founder parents showing contrasting phenotypes for grain yield. We combined phenotypic and genotypic information of this MAGIC population to fit several genomic prediction models which were cross-validated to conduct empirical analyses aimed at examining the predictive ability of these models varying the sizes of training populations. Moreover, several methods to optimize the composition of the training population were also applied to this MAGIC population and cross-validated to estimate the resulting predictive ability. Finally, extensive phenotypic data generated in field trials organized across an ample range of water regimes and climatic conditions in the Mediterranean were used to fit and cross-validate multi-environment genomic prediction models including G×E interaction, using both genomic best linear unbiased prediction and reproducing kernel Hilbert space along with a non-linear Gaussian Kernel. Overall, our empirical analyses showed that genomic prediction models trained with a limited number of MAGIC lines can be used to predict grain yield with values of predictive ability that vary from 0.25 to 0.60 and that beyond QTL mapping and analysis of epistatic effects, MAGIC population might be used to successfully fit genomic prediction models.

We concluded that for grain yield, the single-environment genomic prediction models examined in this study are equivalent in terms of predictive ability while, in general, multi-environment models that explicitly split marker effects in main and environmental-specific effects outperform simpler multi-environment models.

Keywords: genomic prediction, MAGIC, barley, GBLUP, genotype x environment interaction

INTRODUCTION

The experimental design that underlies Multi-parent Advanced Generation Intercrosses (MAGIC) populations traces its origins to the advanced inter-cross lines, which were originally developed in animal model species (Yalcin et al., 2005). MAGIC populations are developed crossing multiple inbred parents or founders, which are subsequently inter-mated several times following pre-defined crossing schemes to shuffle founder genomes in each single line (Huang et al., 2015). In plants, MAGIC populations have been explicitly developed for genetic research purposes as they allow to increase power and precision for detecting and mapping quantitative trait loci (QTLs) (Cavanagh et al., 2008; Huang et al., 2015; Scott et al., 2020). Theoretically, MAGIC populations have the potential to dissect the genetic bases of complex traits at sub-centimorgan scale, allowing to overcome common issues related to the use of biparental families for QTL mapping and detection such as low-resolution power, low genetic diversity of parents and limited number of recombination events (Valdar et al., 2006). In cereal crops, MAGIC populations have been developed and established for rice (Bandillo et al., 2013; Ponce et al., 2018), bread wheat (Mackay et al., 2014; Sannemann et al., 2018; Stadlmeier et al., 2018), maize (Dell'Acqua et al., 2015; Jiménez-Galindo et al., 2019) and barley (Mathew et al., 2018) and to date they have been deployed for unraveling the genetic bases of biotic and abiotic stresses, grain yield (GY) and seed quality traits. Beyond the aforementioned applications, barley MAGIC populations have been recently exploited to disentangle the effect of epistasis on flowering time (Mathew et al., 2018; Sannemann et al., 2018; Afsharyan et al., 2020).

Similarly to MAGIC, the theory underlying genomic prediction (GP) was originally developed and deployed in animal species. The pivotal component of GP is a population of individuals having phenotypic and genotypic information, which is known as training population (TP) and is used to regress genome-wide single nucleotide polymorphisms (SNPs) or other types of DNA markers on phenotypes to simultaneously predict their effects (Meuwissen et al., 2001), that is for training GP models. Trained GP models are subsequently used in combination with the genotypic information of candidate individuals that must be selected for computing their genomic

estimated breeding values (GEBVs) and ranking them to apply truncation selection (Meuwissen et al., 2001; Heffner et al., 2009). This latter population of candidate individuals having only genotypic information is known as breeding population (BP) (Meuwissen et al., 2001; Heffner et al., 2009). To date, GP has been largely applied for crop improvement fitting GP models trained with individuals from either biparental families or diversity panels of mostly unrelated accessions. As the genetic relatedness of TP and BP affects the prediction ability of GP models (Ben Hassen et al., 2018; Norman et al., 2018), these two approaches have profound differences in terms of versatility as DNA marker effects estimated on diversity panels have the potential of a broader applicability and might be used in different breeding programs (Bassi et al., 2015), while GP models trained with individuals of biparental families can allow to accurately predict the performance of offspring produced within the same cross.

Typically, GP models require to regress a number of predictors (DNA markers) that greatly exceeds the number of observations or phenotypes and several parametric and non-parametric models have been proposed to deal with overfitting and the “large p , small n ” problem (Meuwissen et al., 2001; Jannink et al., 2010; Pérez and de los Campos, 2014) as in these conditions the estimation of marker effects using ordinary least squares method is not practicable. A commonly used solution is to estimate marker effects jointly using the Least Absolute Shrinkage and Selection Operator (LASSO) method (Tishbirani, 1996) and its Bayesian counterpart (Bayesian Lasso or BL), which uses a penalizing or regularization parameter (λ) that denotes the amount of shrinkage for regressing markers (De Los Campos et al., 2009). Other popular whole genome regression methods based on Bayesian theory are BayesA and BayesB (Meuwissen et al., 2001), which relax the assumption of common variance across marker effects adopted in other models (e.g., ridge regression) and allow each marker to have its own variance. Differently to BayesA, BayesB allows having markers with no effects in the model and theoretically assumes more realistic conditions as it is plausible that a large fraction of genome-wide markers does not contribute to explaining the observed phenotypic variance. Beyond these methods, whole genome regression based on reproducing kernel Hilbert space (RKHS) has been proposed and applied to implement GP models (Gianola and Van Kaam, 2008; Gota and Gianola, 2014). In the RKHS regression, a reproducing kernel, that is any positive definite function for mapping from pairs of points in input space to other pairs of points, is used to transform DNA markers of individuals in square distance matrix that are used in a linear model (Gota and Gianola, 2014). The Gaussian Kernel (GK) is one of the most common function used as reproducing kernel and depends

Abbreviations: DH, Days to Heading; PH, Plant Height; GY, Grain Yield; MAGIC, Multi-parent Advanced Generation Intercrosses; TP, Training Population; BP, Breeding Population; GP, Genomic Prediction; GBLUP, Genomic Best Linear Unbiased Prediction; RKHS, Reproducing Kernel Hilbert Space; GK, Gaussian Kernel; LD, Linkage Disequilibrium; PCA, Principal Component Analysis; QTL, Quantitative Trait Locus; GEBV, Genomic Estimated Breeding Value; SNP, Single Nucleotide Polymorphism; SE-GP, Single Environment Genomic Prediction; ME-GP, Multi Environment Genomic Prediction.

on the bandwidth (or smoothing) parameter h that controls the decay rate of the kernel as two points step away. Several studies have shown that the use of GK in combination with RKHS improves the prediction of genetic values if the bandwidth parameter h is correctly chosen (Pérez-Elizalde et al., 2015). Moreover as RKHS regression does not assume linearity, this model might allow to better capture non-additive effects without explicitly including epistatic interactions and dominance in GP models (Gianola and Van Kaam, 2008). Differently from methods based on whole genome regression of markers, the genomic best linear unbiased prediction (GBLUP) method treats genomic values of individuals as random effects in a linear mixed model and uses a genomic relationship matrix based on DNA marker data to compute GEBVs (VanRaden, 2008; Wang et al., 2018). Notably, the use of RKHS along with the genomic relationship matrix is equivalent to the mixed linear model of GBLUP, that is GBLUP method represents a special case of RKHS regression (Gota and Gianola, 2014).

The effectiveness of GP depends, among other factors, on the degree of correlation between GEBVs and true genetic values that is the predictive ability of the model. In practice, the predictive ability is evaluated using the Pearson's correlation coefficient between GEBVs and the realized phenotypes or other estimators (e.g., adjusted means). To date several empirical studies have been conducted for fitting GP models on biparental populations and panels of mostly unrelated accessions across different species and traits, which point out that, depending on the genetic architecture of the trait, each statistical model has its own advantages and disadvantages in term of predictive ability and estimation of marker effects (Heslot et al., 2012; Ben Hassen et al., 2018). Other factors that strongly influence the predictive ability are the size of the TP, its structure, and its relatedness with the BP (Desta and Ortiz, 2014). Several targeted and untargeted methods have been developed to optimize the composition of TP for maximizing the predictive ability for a given set of individuals (Rincent et al., 2012; Akdemir et al., 2015). Nevertheless, these methods generally generate trait-dependent TPs which might hamper the implementation of these procedures in real breeding programs.

The first objective of the present study was to create a new barley MAGIC population using a diverse founder set of old and new 6-rowed, winter cultivars showing contrasting GY, which was examined across an ample range of site-by-season combinations characterized by different temperature and precipitation patterns. The second objective of this study was to combine data collected across these field trials with genotypic information to fit different single-environment genomic prediction (SE-GP) and multi environment genomic prediction (ME-GP) models for empirically assessing the predictive ability in multi-parent populations. Moreover, we applied different untargeted optimization methods to this MAGIC population for assembling and benchmarking the performance of optimized TPs. Fitting SE-GP and ME-GP models to MAGIC lines, we aimed at broadening the use of these experimental populations beyond classical QTL mapping and analysis of epistatic effects for sustaining and accelerating barley breeding.

MATERIALS AND METHODS

Development of the Barley MAGIC Population

The MAGIC population used in this study was developed using a founder set of eight 6-rowed barley genotypes with a winter growth habit, which were selected on the basis of their pedigrees and similarity in days-to-heading (DH) (Table 1). At the first stage of MAGIC development, four F_1 populations were created crossing one of the four old 6-rowed barley varieties (Hatif de Grignon, Dea, Robur and Athene) with one of the four 6-rowed modern barley varieties (Ponente, Ketos, Aldebaran and Fridericus). At the second stage of MAGIC development, half-diallel crosses of these four F_1 individuals were carried out to generate six sets of plants. Finally, these six sets of genotypes, each of which contained the alleles of four out eight founder parents, were appropriately crossed in predefined funnel schemes to combine the genome of the eight founders in single lines. Differently from the original crossing schemes developed for constructing MAGIC populations (Cavanagh et al., 2008), instead of recursively self-fertilizing these plants for several generations, seeds of the eight-way inter-crosses were sent to an external lab (SAATEN-UNION GmbH, Germany) to generate 352 inbred MAGIC lines using doubled haploid technology.

Field Trials and Plant Phenotyping

The MAGIC population of 352 inbred individuals and the eight founder parents (Table 1) were sown during the fall of two consecutive growing seasons (2015–2016 and 2016–2017) in Fiorenzuola d'Arda (Italy) at CREA-Centro di Genomica e Bioinformatica (44°55'39.0"N 9°53'40.6"E, 78 m above sea level), using an alpha-lattice design with two-replicates. The whole set of MAGIC and the founder parents were also sown during the fall of 2015–2016 growing season in Marchouch (Morocco) at the Experimental station (33°36'43.5" N 6°42'53.0"W, 390 m above sea level) of the "International Center for Agricultural Research in the Dry Areas" using the same experimental design. Similarly, the subset of 82 MAGIC lines included in the optimized TP (TP-Diverse) and the eight founder parents were sown during the fall in 2017–2018 and 2018–2019 growing seasons in Fiorenzuola d'Arda under two different levels of nitrogen fertilization using alpha lattice experimental designs with two replicates. Trials conducted under ideal nitrogen conditions were fertilized with 100 kg/ha of nitrogen applied in two doses: 50 kg/ha were used at the sowing and 50 kg/ha were applied at the stem elongation stage. Field trials conducted under low nitrogen conditions received 50 kg/ha of nitrogen, 25 of which were applied at sowing while the remaining amount was applied at the stem elongation stage. In the growing season 2018–2019, other two field trials were conducted in Konya (Turkey) (37°53'37.9"N 32°37'26.0"E, 1,005 m above sea level) and in Adana (Turkey) (36°59'52.9"N 35°20'28.0"E, 24 m above sea level) to phenotype the optimized TP (TP-Diverse) using the same experimental design. For each trial considered in this study, plots of three square meters and a sowing density of 350 seeds per square meter were adopted, respectively. Local check cultivars were included

TABLE 1 | Founder set of barley varieties that were intermated for creating the barley MAGIC population.

Genotype	Year of release	Country of release	Pedigree	DH (days)	PH (cm)	GY (t/ha)
Hatif de Grignon	1937	France	Selection from French landraces	208.3	95.9	4.1
Dea	1953	Germany	[(Ragusa × Peragis12) × (Heils Franken × Frw.Berg)] × [(Ragusa × Mahnd.Viktoria) (Ragusa × Bolivia)]	212.1	95.3	6.0
Robur	1973	France	Ager × (Hatif de Grignon × Ares)	208.3	78.8	6.3
Athene	1977	Germany	(Herfodia × Hord.sp.nigrum H204) × (Madru × Weissenhaus-Stamm)	211.5	94.0	6.0
Ponente	2001	Italy	(Vetulkio × Arma) × Express	209.7	85.0	6.3
Ketos	2002	France	(Gotic × Orblonde) × (12813 × 91H595)	208.6	81.9	6.8
Aldebaran	2003	Italy	Rebelle × Jaidor	208.5	83.0	7.2
Fridericus	2006	Germany	Carola × LP 6–564	211.6	89.3	7.3

For each genotype of the founder set, the adjusted means of days to heading (DH), plant height (PH) and grain yield (GY) scored in eight different trials were reported along with available pedigree information.

as internal checks in all experiments to compare phenotypes with trait observations collected in past seasons. Common protocols were adopted for each trial to phenotype plant genotypes for GY and DH. Phenotyping of MAGIC lines for GY was conducted as follows: from each plot grains were collected using a combine harvester and the total grain weight recorded in each plot was converted in tons per hectare. DH was measured as the number of days between sowing date and the date of heading stage, which was defined when 50% of the plants in a plot were at Zadoks' 55 growth stage (Zadoks et al., 1974). For each trial, phenotypic data of GY used in GP models were centered by subtracting the overall mean and standardized dividing by the sample standard deviation.

Statistical Models for Computing the Adjusted Means of GY

The adjusted means of GY were computed in each site-by-season combination and across environments including DH as fixed covariate using the approach described in Emrich et al., 2008. The resulting model for computing the adjusted means of GY collected in field trials organized according to alpha-lattice design was:

$$y_{ijk} = 1\mu + \text{Rep}_i + \text{Block}_j(\text{Rep}_i) + \text{Gen}_k + \text{DH}_k + e_{ijk} \quad (1)$$

where y_{ijk} is the response variable, that is the raw GY, μ is the general mean, Rep_i is the effect of the i^{th} replicate, $\text{Block}_j(\text{Rep}_i)$ is the effect of the j^{th} incomplete block within the i^{th} replicate, Gen_k is the random effect of the k^{th} genotype and DH is the effect of “Days-to-heading” covariate measured in each plot. In this model it is supposed that the random effects of Gen_k follow a normal distribution with mean 0 and variance σ_g^2 , that is $\text{Gen}_k \sim \text{NIID}(0, \sigma_g^2)$, and similarly, the residual terms e_{ijk} are normally distributed with mean 0 and variance equals to σ^2 , that is $e_{ijk} \sim \text{NIID}(0, \sigma^2)$. The adjusted GY values obtained predicting the random terms Gen_k from the aforementioned

model were used as phenotypes for training GP models. The linear mixed model reported in Equation 1 was fitted for each site-by-season combination using R 3.6.2 statistical environment and lme4 package (Bates et al., 2015) and variance components of fitted models were used to compute broad sense heritability (H^2) of GY.

Genotyping of Genetic Materials

DNA was extracted from plant leaves using the Macherey Nagel Plant II extraction kit (Macherey Nagel, Dueren, Germany) and analyzed using gel electrophoresis and Quant-iTTM PicoGreenTM dsDNA Assay Kit (ThermoFisher, Grand Island, NY, United States) following manufacturer's instructions to assess quality and concentration, respectively. DNA samples were shipped to a propel-certified service provider (Trait Genetics GmbH, Gatersleben, Germany) and fingerprinted using the Illumina Infinium technology along with the Barley 50 k iSelect SNP Array (Bayer et al., 2017). To update the physical positions of SNP markers interrogated with the Barley 50 k iSelect SNP Array, probe sets used to design this array were mapped against the new reference sequence of barley (Monat et al., 2019). The raw genotyping table was imported in R software using “synbreed” package (Wimmer et al., 2012) to filter out markers with more than 10% of missing data and impute remaining missing data using Beagle 4.1 (Browning and Browning, 2016). 20 random leaf samples from field trials organized in Adana and Marchouch were genotyped using Illumina Infinium technology and Barley 50 k iSelect SNP Array to assess whether mislabelling of genotypes occurred during phenotyping operations and data collection.

Clustering and Linkage Disequilibrium Analyses of the MAGIC Population

Principal component analysis was used to assess the diversity of the whole MAGIC population and was carried on imputed SNP data of the 352 MAGIC lines and the eight founders using ade4

package along with R version 3.6.2 (Thioulouse et al., 2018; R Core Team, 2019), 2018). The first two principal components were used to visualize the dispersion of MAGIC lines in a graph. Linkage disequilibrium between pairs of markers was measured using r^2 (Hill and Robertson, 2008) in the subset of MAGIC genotypes included in the optimized TP and computed using Plink 1.9 software (Purcell et al., 2007; Chang et al., 2015). r^2 values showing p -values above 0.001 were filtered out, while the remaining pairwise r^2 values were imported and examined with a custom script developed for R 3.6.2 (R Core Team, 2019) to compute the mean r^2 in 100 kb windows, which was plotted in R 3.6.2 using ggplot2 package (Wickham, 2016).

Statistical Models Used for Fitting SE-GP

SE-GP models were fitted using BayesA, BayesB and BL models (Tishbirani, 1996; Meuwissen et al., 2001; Park and Casella, 2008). Moreover, RKHS regression models were fitted using a linear GBLUP kernel (GB) and a non-linear GK (Gianola and Van Kaam, 2008; Gota and Gianola, 2014). For the GK, that is $K(x_i, x_{i'}) = e^{-(h*d_{ii'}^2)}$, where $d_{ii'}^2$ points out the squared Euclidean distance between individuals i and i' , the rate of decay imposed by the bandwidth parameter h , was estimated using an empirical Bayesian methodology (Pérez-Elizalde et al., 2015) modifying published R codes (Cuevas et al., 2016).

Statistical Models Used for Fitting ME-GP

Beyond SE-GP models, the adjusted means of GY computed across different site-by-season combinations were fitted to three previously described ME-GP models. Following the model nomenclature reported in Bandeira e Sousa et al. (2017), these three models were indicated in this study as “multi-environment, main genotypic effect” (MM) model (Jarquín et al., 2014; López-Cruz et al., 2015; Bandeira e Sousa et al., 2017), “multi-environment, single variance G×E deviation model” (MDs) (Jarquín et al., 2014; Bandeira e Sousa et al., 2017) and the “multi-environment, environment-specific variance G×E deviation model” (MDe) (López-Cruz et al., 2015; Bandeira e Sousa et al., 2017). Site-by-season combinations were considered as environments in MM, MDs and MDe regression models, which are briefly defined and summarized as follows. In the MM model, environments were considered as fixed effects while the random genetic effects were considered constant across all environments without modeling marker x environment interactions. Following matrix notation, the MM regression model is defined as follows:

$$y = 1\mu + Z_e\beta_e + Z_uu + \varepsilon \quad (2)$$

where y is the vector of observations collected in all environments, 1 is the overall mean, Z_e is the incidence matrix that connects observed phenotypes to the environments in which they were measured, β_e is the vector of environmental fixed effects that must be estimated, Z_u is an incidence matrix connecting genotypes with phenotypes for each environment, u is the vector of random genetic effects that must be predicted while ε is a vector of model residuals. In this model, marker

genetic effects are assumed as $u \sim N(0, \sigma_{\mu 0}^2 K)$, that is, they follow a multivariate normal distribution with mean and variance-covariance matrix equal to zero and $\sigma_{\mu 0}^2 K$, respectively. The term $\sigma_{\mu 0}^2$ of the variance-covariance matrix is the variance of additive genetic effects across environments, while K can be either a genomic relationship matrix (VanRaden, 2008) or a kernel function as discussed below. Model residuals of the vector are assumed to be independent and normally distributed with null mean and variance equal to σ_e^2 , that is $\varepsilon \sim N(0, I\sigma_e^2)$, where I points out the identity matrix. Overall, the MM regression model estimates marker effects across all environments and does not split them in main marker effects and in environmental-specific effects as in MDs and MDe models. As already substantiated in López-Cruz et al. (2015), for balanced field trial designs, MM is equivalent to fitting a genomic regression model using the average performance of each line across environments as phenotype.

Differently from the MM model, the MDe model allows markers to assume different effects in each j^{th} environment (López-Cruz et al., 2015; Bandeira e Sousa et al., 2017), and consequently allows to account for marker x environment interactions. This model assumes that the effects of the j^{th} environments, and the effects of markers are separated into two components, which are the main effect of markers for all environments, names as b_{0k} , and the peculiar random effect b_{ik} , of the markers in each j^{th} environment, that is the effects of marker x environment interactions (López-Cruz et al., 2015). Consequently, in MDe models, the effect of the k^{th} marker on the j^{th} environment (β_{jk}) is described as the sum of an effect common to all environments (b_{0k}), plus a random deviation (b_{ik}) peculiar to the j^{th} environment, that is $\beta_{jk} = b_{0k} + b_{ik}$.

Following matrix notation, the MDe regression model is defined as follows:

$$y = 1\mu + Z_e\beta_e + Z_uu_o + u_E + \varepsilon \quad (3)$$

where, Z_e , β_e have the same meaning of the MM regression model, u_o represents the main effect of markers across all environments with a variance-covariance structure similar to MM model, that is, $u_o \sim N(0, \sigma_{\mu 0}^2 K)$. As pointed out by López-Cruz et al. (2015) $\sigma_{\mu 0}^2$ is common to all environments, and the borrowing of information among environments is generated through the kernel matrix K . u_E points out the specific effects of marker x environment interactions, which follow a multi-variate normal distribution with null mean and a variance-covariance matrix K_E , that is, $u_E \sim N(0, K_E)$. For j environments, the variance-covariance matrix K_E is defined as follows:

$$K_E = \begin{bmatrix} \sigma_{\mu E1}^2 K_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{\mu Em}^2 K_m & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \sigma_{\mu Ej}^2 K_j \end{bmatrix}$$

As explained in Bandeira e Sousa et al. (2017), K_E can be decomposed as a sum of j matrices, one for each j environment.

Consequently, the interaction term u_E can be decomposed in j environmental specific effects to transform equation 3 as follows:

$$y = 1\mu + Z_e\beta_e + Z_uu_0 + u_{E1} + u_{E2} + u_{E3} + \dots + u_{Ej} + \varepsilon \quad (4)$$

where each interaction effect u_{Ej} has a normal distribution with null mean and a variance-covariance structure $\sigma_{\mu_{Ej}}^2 K_j$.

Starting from the MM regression model, the MDs model adds the random interaction effect of the environments with the genetic information of the lines pointed out with u_e . Following matrix notation, the MDs model is described as follows:

$$y = 1\mu + Z_e\beta_e + Z_uu + u_e + \varepsilon \quad (5)$$

where, Z_e , β_e , Z_u , u and ε have the same meaning of the MM regression model. As substantiated in Jarquín et al. (2014) the interaction term u_e has a multi-variate normal distribution with null mean and variance-covariance matrix equal to $[Z_u K Z_u'] \circ [Z_e Z_e']$, where the Haddamar product operator denotes the element to element product between the two matrices in the same order.

In the present study, MM, MDs and MDe regression models were fitted using either the linear GB kernel method (VanRaden, 2008) or the non-linear GK method (Bandeira e Sousa et al., 2017). For the linear GB kernel method, the matrix K of the aforementioned models was the genomic relationship matrix and was computed as $K = \left(\frac{XX'}{p}\right)$ (VanRaden, 2008), where X is the standardized matrix of molecular markers for the individuals, of order n by p ; where n and p are the number of observations and the number of markers, respectively. For GK method, the matrix K of MM, MDs and MDe regression models was computed as $K_j(x_{ij}, x'_{ij}) = e^{-(h_j * d_{ii}^{2'})}$ where $d_{ii}^{2'}$ is the squared Euclidean distance of the markers genotypes in individuals i and i' for the j^{th} environment. Similarly to SE-GP models, the bandwidth parameter h was computed using an empirical Bayes method (Pérez-Elizalde et al., 2015; Cuevas et al., 2016).

MM, MDs and MDe regression models used in this study were fitted using BGLR package 1.08 (Pérez and de los Campos, 2014) in R 3.6.2 statistical environment, adapting scripts provided in the framework of other studies (Bandeira e Sousa et al., 2017). For each model implemented in this study, predictions were based on 500,000 iterations collected after discarding 10,000 iterations for burn-in period and using a thinning interval of five iterations. Trace plots for each of the variance parameters were created to assess whether the number of burn-in iterations was sufficient.

Optimization of the TPs

In this study three different untargeted optimization criteria based on coefficient of determination (Laloe, 1993), predictive error variance (Rincent et al., 2012) and rScore (Ou and Liao, 2019) were used to assemble three corresponding TPs,

each of which groups a set of 90 MAGIC individuals. The R package TSDFGS (Ou and Liao, 2019) was used to assemble these three optimized TPs using the aforementioned criteria. A fourth empirical untargeted optimization criterion was adopted for assembling another TP from the whole MAGIC population and aimed at maximizing the average distance between each selected accession and the closest other line using the modified Roger's distance (Thachuk et al., 2009). This criterion was implemented in R 3.6.2 using the heuristic algorithm implemented in the package Core Hunter3 (De Beukelaer et al., 2018) and was used to select a subset of 82 out of 352 MAGIC individuals along with the eight MAGIC founder parents.

Cross Validation Schemes

In this study several cross-validation (CV) schemes were adopted for estimating the predictive ability of GP models along with their standard errors (Burgueño et al., 2012; Gianola and Schon, 2016). For estimating the predictive ability of SE-GP models implemented with BayesA, BayesB, Bayesian Lasso, GB and RKHS with GK, cross validation was carried out using 100 repeated random partitioning of MAGIC population into training and validation sets. Using increasingly larger TPs of 80, 90, 100, 110, 120, 130, 140, 150, and 160 individuals, CV schemes were applied to compute mean and standard deviation of predictive ability for each TP size. Totally 4,500 models were fitted to carry out this CV experiment, combining the five statistical models with the aforementioned dimensions of the TP and 100 repeated random partitioning of MAGIC in training and validation sets.

Cross-validation of SE-GP models fitted using optimized TPs was carried out using the standard leave-one-out (LOO) strategy to estimate their predictive ability (Gianola and Schon, 2016). Basically, using LOO strategy, N GP models are fitted using $N-1$ individuals excluding recursively one individual from the TP and the GEBV of the excluded line is predicted from a model trained using all other lines. In our LOO experiment, this was carried out separately for each group of 90 lines included in the optimized TPs, and the accuracy of these predictions was calculated as the Pearson's correlation coefficient between GEBVs and the corresponding adjusted means of GY.

The predictive ability of ME-GP models was assessed using cross-validation 1 (CV1) and cross-validation 2 (CV2) schemes (Burgueño et al., 2012), assigning 90% of lines to the training set and the remaining 10% to the validation set. In both CV schemes, all the parameters of the MM, MDs and MDe regression models were recursively re-estimated in each of 100 random partitions. For each random partitioning, models were fitted using genotypes included in the training sets and the predictive ability was computed as the Pearson's correlation coefficient between GEBVs and the corresponding adjusted means of GY. Overall, 100 Pearson's correlations were computed for each model and the mean and standard deviation of these values were computed to estimate the predictive ability of GP models.

RESULTS

Development of the Barley MAGIC Population

The barley genotypes included in the founder set of MAGIC were examined in field trials organized in height site-by-season combinations in Italy, Germany and Scotland (Xu et al., 2018) for assessing the diversity of European cultivars for GY, plant height and DH. These field trials showed that the founder set, which includes four elite and four old barley varieties with different genetic background, exhibits limited variation of DH values (Table 1). Following a modified version of the standard crossing design (Huang et al., 2015), this founder set was intermated to create an eight-way MAGIC population of 352 individuals, which were subsequently genotyped to assess the contribution of each founder parent to the mosaic genome of each line.

Estimating the Predictive Ability of GP Models as a Function of TP Size

In GP models, the variation of predictive ability as a function of the TP size has been empirically investigated on segregating families and in collections of mostly unrelated accessions (Norman et al., 2018). Here, we investigated the relationship between TP size and the predictive ability of different GP statistical models fitted to the barley MAGIC population. To carry out this analysis, the whole panel of 352 MAGIC lines and the founder parents were genotyped using the Barley 50 k iSelect SNP Array (Bayer et al., 2017). SNPs with more than 10% of missing data were discarded, while the remaining missing genotypes were imputed using the algorithm implemented in BEAGLE (Browning and Browning, 2016). This procedure allowed to identify 19,723 polymorphic SNPs, which were combined to the adjusted means (BLUPs) of GY computed in three site-by-season combinations (Table 2) to fit and cross-validate SE-GP models. Overall, five different whole genome regression methods based on BayesA, BayesB, BL, GB and RKHS fitted with the non-linear GK (Gianola and Van Kaam, 2008; Gota and Gianola, 2014; Cuevas et al., 2016; Crossa et al., 2017) were compared.

These aforementioned SE-GP models were fitted to the MAGIC population and cross-validated for estimating the trend of predictive ability as a function of TP size (Figure 1). Specifically, CV was implemented randomly partitioning 100 times the whole panel of MAGIC lines in a TP and in a validating population (VP). Overall, nine different CV experiments were carried out, using TP sizes of 80, 90, 100, 110, 120, 130, 140, 150, and 160 MAGIC lines and the remaining genotypes as VPs (Figure 1). The CV of these GP models points out that in the three site-by-season combinations (Table 2), GB, GK, BayesA, BayesB and BL show comparable predictive abilities across the entire range of TP sizes considered (Figure 1). Moreover, these CV experiments point out that in temperate locations (Fio16IN, Fio17IN, Table 2), the predictive ability of SE-GP models exceeds 0.50 even using TPs of 80 or 90 individuals (Figure 1), while in the harsh and pre-desertic environment of Mar16IN (Table 2), it does not exceed 0.25 and shows larger standard deviation. Varying the

size of TPs from 80 to 160 individuals slightly increases the values of predictive ability for GY in the remaining individuals of the MAGIC population (Figure 1 and Supplementary Table 1) as already substantiated in other GP models fitted using collection of mostly unrelated genotypes (Norman et al., 2018). Overall, this empirical analysis shows that 80 or 90 MAGIC individuals are sufficient to fit SE-GP models yielding high values of predictive ability and that larger TPs do not significantly improve the predictive ability of GP models either in temperate or stressful environments (Figure 1 and Supplementary Table 1).

Designing Optimized TPs of MAGIC

The predictive ability of GP models fitted in collection of mostly unrelated accessions and in biparental populations depends on the size of TP, the genome distribution and number of molecular markers used for whole genome regression, the genetic composition of TP and its genetic relationship with the BP (Heffner et al., 2009; Jannink et al., 2010; Desta and Ortiz, 2014; Berro et al., 2019). Particularly, it was assessed that using a large reference panel of accessions, the predictive ability of GP models can be improved increasing the diversity of the TPs (Norman et al., 2018). Along with these empirical findings, several statistical criteria and algorithms have been proposed to optimize TPs for maximizing predictive ability using reference panels of accessions or sets of advanced lines (Akdemir et al., 2015; Berro et al., 2019; Ou and Liao, 2019).

Here, we examined three different untargeted optimization criteria based on the coefficient of determination (CD_mean) (Laloe, 1993), prediction error variance (PEV) (Rincent et al., 2012) and rScore (Ou and Liao, 2019) and benchmarked them against a method that samples a diverse TP from the whole MAGIC population using SNP markers (Figure 2). The rationale of this latter method is to maximize the average distance, computed using the modified Roger's method, between each selected accession and the closest other genotype (Thachuk et al., 2009). This criterion, named entry-to-nearest entry was maximized with a heuristic algorithm to construct a highly diverse TP in which all MAGIC lines are maximally different (De Beukelaer et al., 2018). The TP assembled with this latter untargeted optimization criterion, named "TP-Diverse" (Figure 2), was constructed using the panel of 19,723 polymorphic SNPs detected in the whole MAGIC population, and was subsequently used as optimized TP and benchmarked to TPs assembled using CD_mean, PEV and rScore optimization methods (Figure 2).

Following this "TP-Diverse" optimization, our procedure led to identify a set of 82 MAGIC lines as the smallest population subset fulfilling the aforementioned criterion, which was used as TP along with the eight founder parents. Overall, when applied to MAGIC populations, the four optimized TPs spawned similar predictive abilities across the three site-by-season combinations (Figure 2) and consequently the genetic makeup of this TP was further investigated. The genetic relationships between TP-Diverse and the remaining MAGIC lines was assessed conducting a principal component analysis (PCA) on genetic data, which pointed out that the first two principal components explain 22.3 and 5.5 percent of the total genetic variability of the

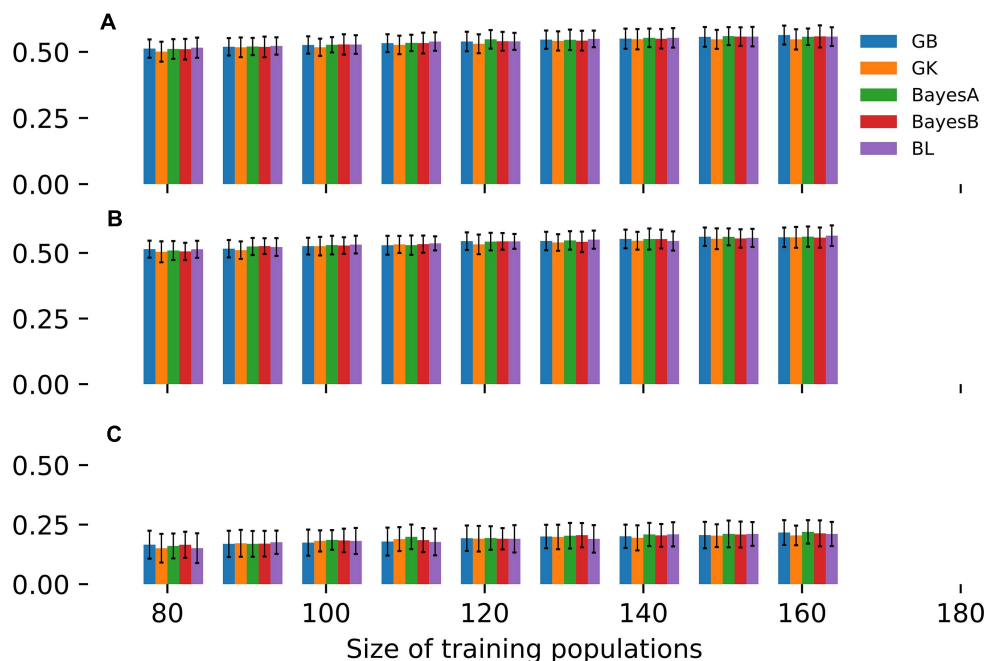


FIGURE 1 | CV of different SE-GP models fitted to GY measured in the MAGIC population. Bars report the values of predictive ability for GY computed in (A) Fio16IN, (B) Fio17IN, and (C) Mar16IN. Bars of different colors point out values of predictive ability computed using GB, GK, BayesA, BayesB and BL models as a function of TP sizes, while the error bars point out the standard deviation of predictive ability values.

TABLE 2 | Field trials carried out for phenotyping the whole MAGIC population and the founder set for GY.

Acronym	Site	Country	Growing season	Populations	Traits
Fio16IN	Fiorenzuola d'Arda	Italy	2015–2016	352 MAGIC and the founder set	DH, GY
Fio17IN	Fiorenzuola d'Arda	Italy	2016–2017	352 MAGIC and the founder set	DH, GY
Mar16IN	Marchouch	Morocco	2015–2016	352 MAGIC and founder set	DH, GY

MAGIC population, respectively (**Figure 3**). PCA shows three main clusters of MAGIC lines and corroborates that individuals included in the TP-Diverse are representative of the whole diversity of MAGIC lines (red points).

In segregating families and collections of mostly unrelated accessions, a large number of molecular markers is often needed to capture the effects of all QTLs or alternatively, strong linkage disequilibrium (LD) between markers and causative variants that control the traits of interest is desirable to achieve high values of predictive ability in GP (Lorenzana and Bernardo, 2009; Heffner et al., 2011; Norman et al., 2018). Consequently, the extent of LD was investigated in TP-Diverse to assess its correlation with the predictive ability values of GP models. Firstly, SNP markers of the barley 50 K SNP chip used to fingerprint the whole MAGIC population were lifted over to the new barley reference sequence (Monat et al., 2019) and secondly, the average extent of r^2 was computed for each barley chromosome. Overall, a large fraction of the 44,040 SNPs of the barley 50 k SNP chip were lifted over and 18,248 out 19,723 polymorphic SNPs unambiguously mapped to the reference sequence of barley (**Supplementary Table 2**) were used to estimate the decay of average LD computed in bins of 100 kb (**Figure 4**). This analysis indicated that across

the seven barley chromosomes r^2 decays relatively slowly as SNPs mapped more than 10 Mbp apart show r^2 values of circa 0.2, while the average r^2 values of markers within 1 MB or less exceed 0.4 (**Figure 4**). Considering the average number of markers per chromosome (**Supplementary Table 2**), the levels of LD measured in TP-Diverse are sufficiently high and higher marker densities might not significantly increase the predictive ability of GP models fitted in our MAGIC population of barley as empirically observed in other crops (Norman et al., 2018). Overall, the predictive ability values obtained with GP models fitted with the three optimization methods are substantially equivalent to the prediction accuracy obtained with TP-Diverse (**Figure 2**) and consequently this latter TP was chosen for fitting further single- and multi-environment GP models.

Using the Optimized TP for Fitting SE-GP and ME-GP Models

Field trials of TP-Diverse were organized in nine site-by-season combinations and phenotypic data for GY and DH were collected using common phenotyping protocols, while the remaining set of MAGIC lines were used in Fio16IN, Fio17IN

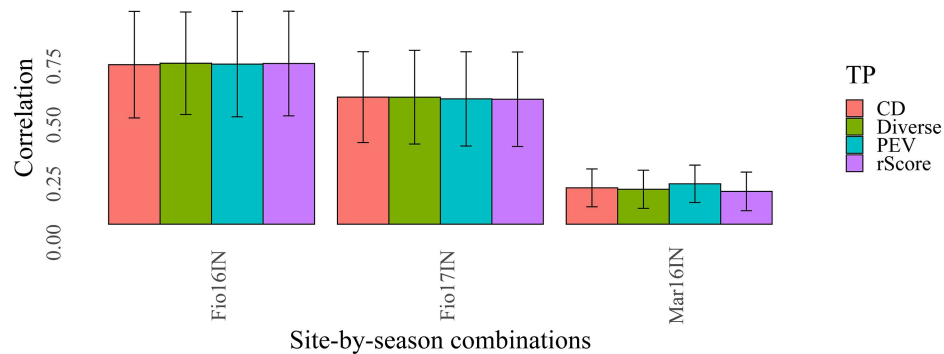


FIGURE 2 | Benchmarking of different methods for optimizing TPs of MAGIC. Bars of different colors report the values of predictive ability obtained with GP models fitted using CD_mean (CD), prediction error variance (PEV), rScore and Diverse optimization criteria. The error bars of each plot point out the standard deviation of the predictive ability values.

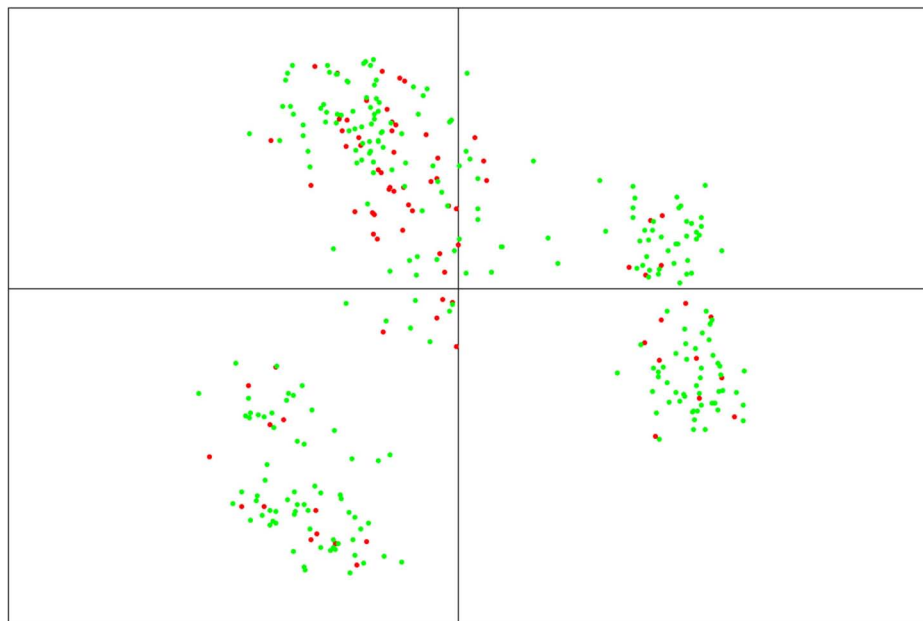


FIGURE 3 | Principal component analysis (PCA) of the MAGIC population based on 19,723 SNPs. The first two axes of PCA explain 22.3 and 5.5% of the total variability, respectively. Red points represent the subset of MAGIC lines included in TP-Diverse, while green points represent the remaining MAGIC lines.

and Mar16IN as VP (**Table 3**). Alpha-lattice experimental designs were adopted for all field trials and mixed linear models were used to compute adjusted means of GY and broad sense heritability (H^2) for each site-by-season combination considering genotypes as random variables (BLUPs) (**Table 3**). This analysis indicated that H^2 varies significantly across the nine field trials and spans from 0.805 in Kon19IN to 0.122 in Mar16IN (**Table 3**). The adjusted means of GY were subsequently used as phenotypes for fitting GP models along with genotypic information.

To assess the performance of MAGIC lines included in TP-Diverse, across different locations and years, a pairwise correlation analysis of the adjusted means of GY computed in the nine site-by-season combinations considered in this

study was carried out (**Figure 5**). The correlations of GY across environments spanned from -0.030 to 0.553 and, as expected, values were higher between field trials carried out in the same environments but in different years, while lower values were observed among Mar16IN and other site-by-season combinations, corroborating the hypothesis that the climatic peculiarity of this environment imposes higher levels of stress to MAGIC lines (**Figure 5**). Similarly, the adjusted means of GY computed in Fio18LN exhibited lower correlation values with other site-by-season combinations (**Figure 5**). These adjusted means of GY were used to train SE-GP and ME-GP models using “TP-Diverse.” For each site-by-season combination, phenotypic and genotypic data were standardized, and nine different SE-GP models were fitted using GB and GK statistical models (**Table 4**).

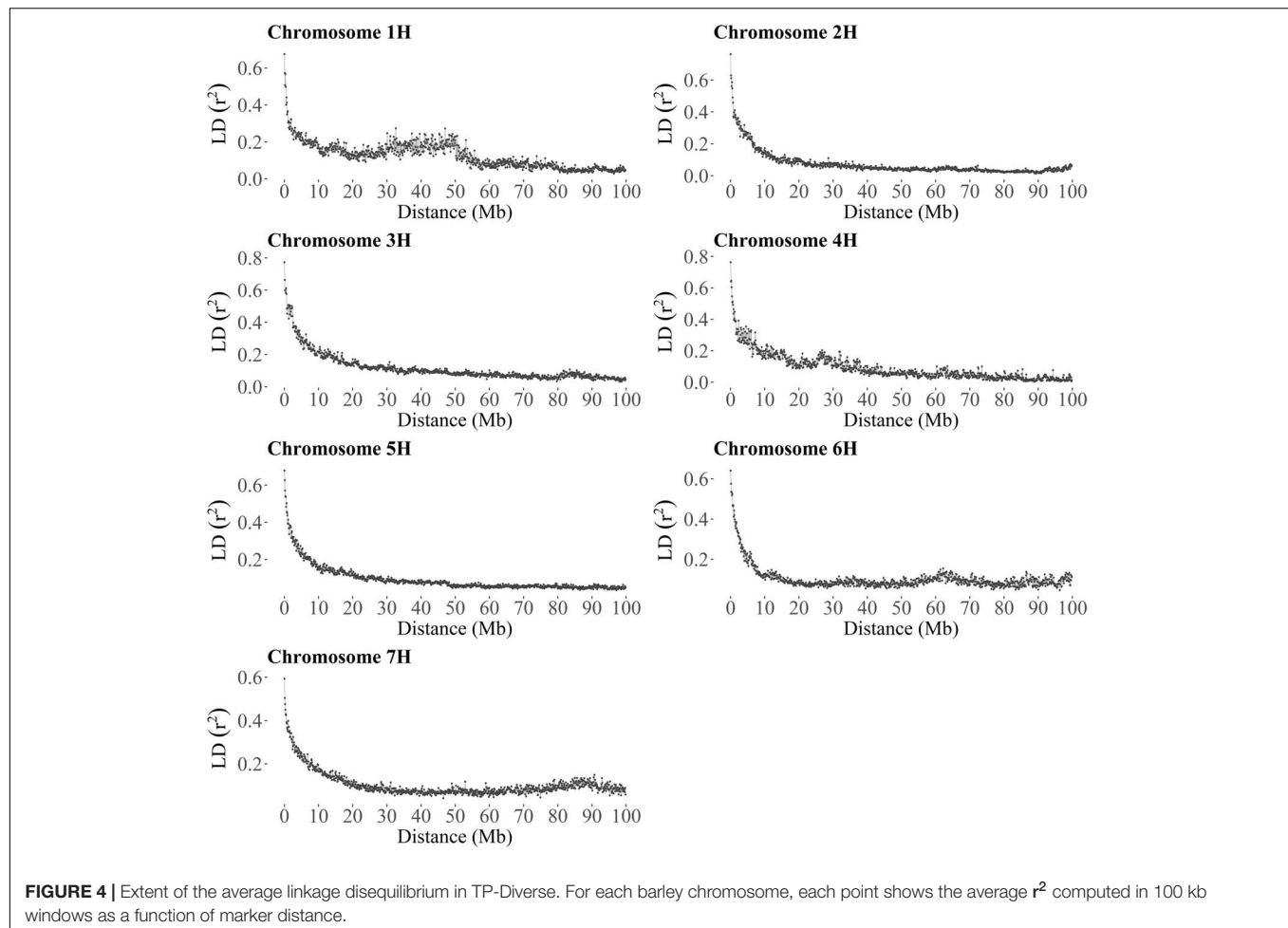


TABLE 3 | Summary of field trials carried out for phenotyping TP and VP for GY.

Acronym	Site	Country	Growing season	Populations	H ²
Fio16IN	Fiorenzuola d'Arda	Italy	2015–2016	TP and VP	0.660
Fio17IN	Fiorenzuola d'Arda	Italy	2016–2017	TP and VP	0.472
Fio18IN	Fiorenzuola d'Arda	Italy	2017–2018	TP	0.532
Fio18LN	Fiorenzuola d'Arda–Low Nitrogen	Italy	2017–2018	TP	0.395
Fio19IN	Fiorenzuola d'Arda	Italy	2018–2019	TP	0.652
Fio19LN	Fiorenzuola d'Arda–Low Nitrogen	Italy	2018–2019	TP	0.663
Mar16IN	Marchouch	Morocco	2015–2016	TP and VP	0.122
Ada19IN	Adana	Turkey	2018–2019	TP	0.737
Kon19IN	Konya	Turkey	2018–2019	TP	0.805

For each site-by-season combination, the estimates of broad sense heritability (H^2) of GY were reported. H^2 was computed for the whole panel of MAGIC lines for Fio16IN, Fio17IN and Mar16IN.

As expected after standardization, for models fitted using GB, the summation of variance components was circa 1 (Table 4), while the distribution of the residuals after fitting all GP models to the nine site-by-season combinations was approximately normal. The analysis of variance components of SE-GP models showed that the values of error variance in GK models are lower than those obtained for the corresponding GB models (Table 4), and similarly in GK models the values of genetic component variance

are always higher than the corresponding quantities computed for GB models (Table 4).

The adjusted means of GY computed at the nine site-by-season combinations were used to fit ME-GP, particularly three models were fitted, which were named “Multi-environment, main genotypic effect” (MM), “Multi-environment, single variance GxE deviation” (MDs) (Jarquín et al., 2014) and “Multi-environment, environment specific variance GxE deviation”

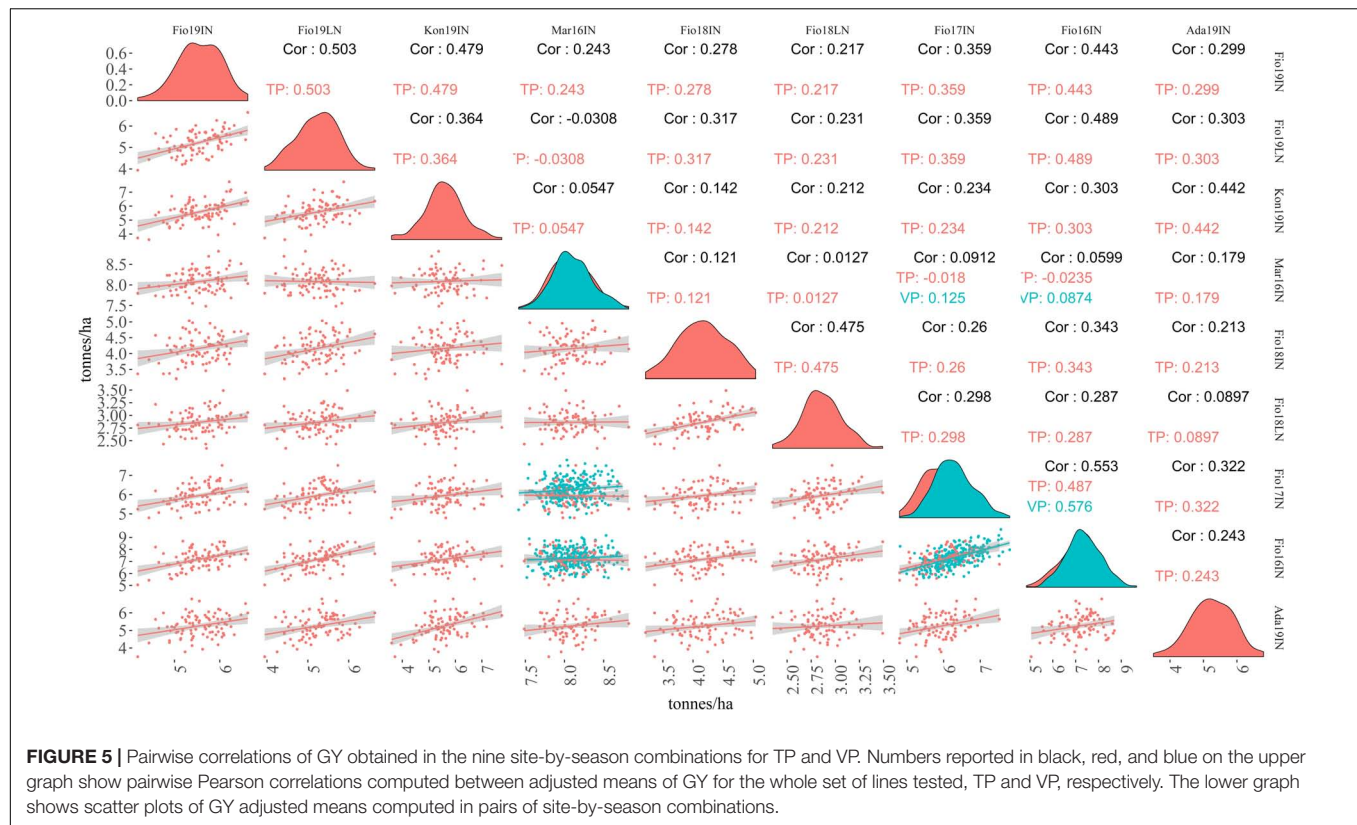


TABLE 4 | Variance components of SE-GP models fitted using GBLUP (GB) and GK statistical model.

Site-by-season combination	GB		GK	
	Genetic effect variance	Residual variance	Genetic effect variance	Residual variance
Kon19IN	0.586 (0.010)	0.557 (0.045)	0.660 (0.016)	0.489 (0.068)
Mar16IN	0.467 (0.089)	0.719 (0.067)	0.590 (0.013)	0.588 (0.078)
Fio18IN	0.491 (0.059)	0.560 (0.029)	0.632 (0.086)	0.455 (0.043)
Fio18LN	0.412 (0.048)	0.752 (0.050)	0.544 (0.000)	0.611 (0.069)
Fio17IN	0.537 (0.072)	0.480 (0.016)	0.655 (0.084)	0.417 (0.041)
Fio16IN	0.618 (0.066)	0.336 (0.094)	0.680 (0.065)	0.348 (0.011)
Ada19IN	0.561 (0.086)	0.543 (0.036)	0.654 (0.019)	0.498 (0.070)
Fio19IN	0.480 (0.079)	0.651 (0.049)	0.659 (0.005)	0.469 (0.054)
Fio19LN	0.479 (0.058)	0.566 (0.024)	0.632 (0.083)	0.446 (0.041)

For each site-by-season combination, the estimated variance components of genetic effects and residuals fitted with GB and GK models are reported, while bracketed numbers point out the corresponding standard deviation.

(MDe) (López-Cruz et al., 2015) following recent model nomenclature (Bandeira e Sousa et al., 2017). Similarly to SE-GP models, MM, MDs, and MDe models were fitted using GB and GK methods and totally six model method combinations were used to fit multi-environment predictions. The analysis of variance components showed that for all three models (MM, MDs, and MDe), GK methods exhibit lower values of the estimated residual variances pointing out a better model fitting (Table 5). Moreover, model comparisons showed that the inclusion of the interaction term (GxE) in MDe model induces a reduction in the estimated residual variance for GY compared to MM models either using GB or GK methods, but

MDs models fitted better the data compared to MDe. For the MDe models, the residual variance components of MDe-GK were smaller than those of the MDe-GB, whereas the estimated variance components for the genetic main effect and genetic environment specific effect variances were higher for the GK than for the GB (Table 5).

Predictive Ability of ME-GP Models With GB and GK Methods

The predictive ability of MM, MDs, and MDe models implemented using GB and GK methods was estimated

with cross-validation 1 (CV1) and cross-validation 2 (CV2) schemes using 100 random partitions. For each of the six multi-environment model-method combinations, the values of predictive ability for CV1 and CV2 schemes were obtained for the set of 100 random partitions, which were used to compute the average predictive ability and the associated standard deviation. Overall, CV2 showed that in four site-by-season combinations (Fio16IN, Fio17IN, Fio19IN, and Fio19LN) the predictive ability is generally higher and exceed 0.70 for certain ME-GP models, while for Mar16IN the six model-method combinations exhibit, on average, the lowest values of predictive ability as for this site-by-season combination the lowest values of 0.161 and 0.236 were observed for MM-GB and MDs-GK models, respectively (Figure 6 and Supplementary Table 3).

As in most of the case, the standard deviations associated to the values of predictive ability were overlapping (Figures 6, 7), Welch's *t*-tests were applied to determine whether pairwise comparisons of predictive ability values obtained with ME-GP models were statistically different (Supplementary Figures 1, 2). CV2 experiments showed that in Fio17IN the values of predictive ability computed with the six-model method combinations were comparable except for MM-GB, which was significantly lower than the predictive ability of MDs-GK, while in Fio16IN the

predictive ability of MM-GK was significantly lower than the predictive ability obtained with the remaining model-method combinations (Figure 6 and Supplementary Figure 2). In Fio16IN, CV2 showed that MDe-GB and MDe-GK have similar performance and significantly higher values of predictive ability compared to MM models, either implemented with GB or GK statistical methods (Figure 6, Supplementary Table 3, and Supplementary Figure 2). In Ada19IN the best model predictive ability using CV2 scheme was obtained with MDe-GB, while for Fio18LN the best values of predictive ability were obtained with MDe-GB and MDs-GB models. Overall, CV2 experiments indicated that in four out nine site-by-season combinations (Fio16IN, Fio17IN, Fio18IN, and Mar16IN) MDe-GB and MDe-GK models have higher values of predictive ability compared to MM models, either implemented with GB or GK statistical methods (Figure 6, Supplementary Table 3, and Supplementary Figure 2). Differently, Fio19IN, Fio19LN, and Kon19IN deviate from this trend as for these site-by-season combinations the values of predictive ability for MM models were higher (Supplementary Table 3). In Fio19IN, MM-GB and MM-GK had the higher predictive ability values along with MDe-GK, while for Fio19LN the higher value of predictive ability was found for MM-GB.

The values of predictive ability obtained for random CV1 decreased (Figure 7 and Supplementary Table 4) as compared with those computed for CV2 for all models. Similarly to the results obtained for CV2, CV1 experiments indicated that in four site-by-season combinations (Fio16IN, Fio17IN, Fio18IN, and Fio19LN) the predictive ability of GP-ME models is generally higher than the values of predictive ability observed in other site-by-season combinations for all models. MDs-GB and MD-GK yielded the higher values of predictive ability in Ada19IN, Fio16IN, and Fio17IN, respectively. In Fio18IN, Fio18LN, Mar16IN, and Fio19LN, the higher predictive ability values were found for MM-GK, although in this latter site-by-season combination the accuracy of MDe-GK does not differ significantly (Supplementary Figure 1). In Fio19IN, the highest values of predictive ability were obtained for MDe-GB and MD-GK models (Figure 7 and Supplementary Figure 1).

TABLE 5 | Variance components of ME-GP models fitted using GBLUP (GB) and RKHS along with the Gaussian Kernel (GK) methods.

Component	Environment	GB	GK
Multi-environment, main genotypic effect (MM) model			
Residual (σ_e^2)	–	0.758 (0.047)	0.746 (0.045)
Genetic main effect ($\sigma_{\mu_0}^2$)	–	0.249 (0.069)	0.373 (0.088)
Multi-environment, single variance GxE deviation (MDs) model			
Residual (σ_e^2)	–	0.516 (0.056)	0.389 (0.071)
Genetic main effect ($\sigma_{\mu_0}^2$)	–	0.281 (0.077)	0.374 (0.089)
Genetic interaction effect ($\sigma_{\mu_{ge}}^2$)	–	0.247 (0.066)	0.589 (0.140)
Multi-environment, environment specific variance GxE deviation (MDe) model			
Residual (σ_e^2)	–	0.602 (0.016)	0.592 (0.018)
Genetic main effect ($\sigma_{\mu_0}^2$)	–	0.292 (0.026)	0.402 (0.031)
Genetic environment specific effect ($\sigma_{\mu_{Ej}}^2$)	Ada19IN	0.251 (0.054)	0.353 (0.083)
	Fio16IN	0.035 (0.027)	0.054 (0.046)
	Fio17IN	0.010 (0.066)	0.024 (0.023)
	Fio18LN	0.062 (0.054)	0.116 (0.085)
	Fio18IN	0.007 (0.006)	0.018 (0.015)
	Mar16IN	0.549 (0.085)	0.873 (0.122)
	Kon19IN	0.217 (0.050)	0.312 (0.079)
	Fio19LN	0.008 (0.007)	0.053 (0.018)
	Fio19IN	0.004 (0.003)	0.055 (0.011)

For each of the three regression models (MM, MDs and MDe), the estimated variance components fitted with GB and GK methods are reported, while bracketed numbers point out the corresponding standard deviation of variance component estimates.

DISCUSSION

Broadening the Use of MAGIC Populations for Plant Breeding

Multi-parent Advanced Generation Intercrosses populations were conceived to improve precision and efficiency of QTL mapping in plants and animals as they allow overcoming limitations of biparental populations and association mapping panels (Huang et al., 2015). In cereal crops, these experimental populations have been extensively used for research purpose and contributed to dissecting the genetic bases of several traits among which biotic stress resistance (Stadlmeier et al., 2018; Jiménez-Galindo et al., 2019; Riaz et al., 2020), GY, grain quality (Zaw et al., 2019) and DH (Afsharyan et al., 2020). Recently, these genomic resources have been established in barley to investigate the effects of epistasis and environmental interactions on flowering time

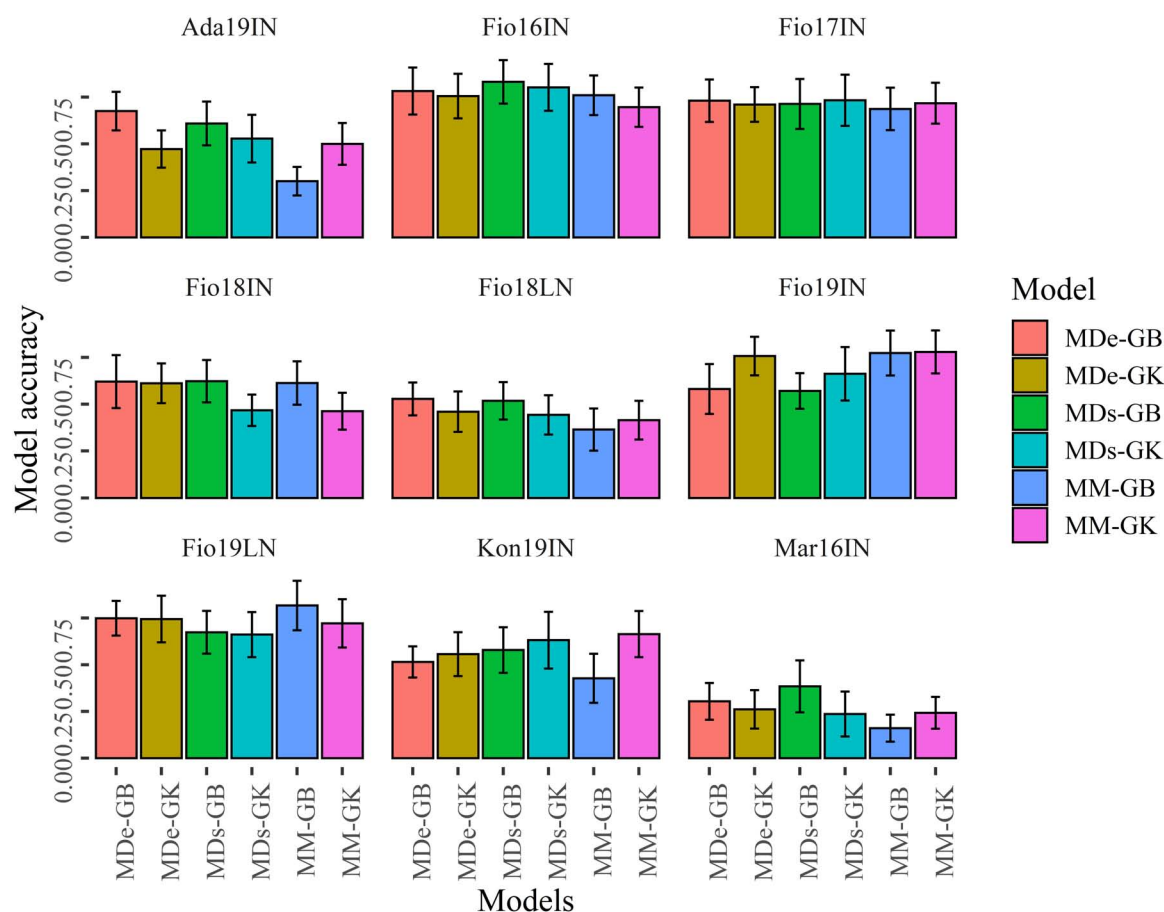


FIGURE 6 | Bar plots of the predictive ability values obtained with CV2. Bar plots show the mean correlation between observed and predicted values of GY obtained with 100 random CV2 partitions for MM, MDs and MDe models implemented with GBLUP (GB) and Gaussian Kernel (GK) methods. Error bars point out the standard deviation of predictive ability values.

(Mathew et al., 2018; Afsharyan et al., 2020), further broadening the original scope for which they were devised.

In the present study, we constructed a new MAGIC population shuffling alleles of winter 6-rowed barley varieties, and demonstrated that, along with biparental populations and collections of mostly unrelated accessions, these genomic resources might be used to train GP models with high predictive ability and might speed up barley breeding. Under this point of view, the large number of MAGIC populations developed in the last years in several crops (Kover et al., 2009; Rebetzke et al., 2014; Mathew et al., 2018; Stadlmeier et al., 2018) can be considered as untapped resources that would contribute to further strengthening and stimulating the application of GP in plant breeding. On the other side, *de novo* creation of MAGIC populations to train GP models for actual breeding purposes is hampered because of their time consuming and costly development, which requires to intermate and self-fertilize the founder parents for several cycles. The results presented in this study show that these limitations might be softened using doubled haploid technology, which allows to short self-fertilization stages to obtain fully homozygous lines. Similarly,

speed breeding might contribute to accelerating the development of new MAGIC populations (Watson et al., 2018).

To examine the genetic relationship between the whole set of MAGIC and the subset of lines included in the “TP-Diverse,” a PCA was carried out using 19,723 SNPs, which detected genetic structure in the MAGIC population and three main clusters of individuals. The nature of these clusters is unclear, but it is plausible that they might reflect subgroups of individuals showing segregation distortion for one or more founders. In our eight-way MAGIC population, the expected segregation rate of the eight founder haplotypes is 1:1:1:1:1:1:1:1, but the haplotypes of some founders (e.g., Dea) deviate from the expected ratio (**Data not shown**). Segregation distortion is a common phenomenon that occurs in MAGIC populations as pointed out in other studies (Sannemann et al., 2018). Although this did not hamper our ability to train GP models with this population, this phenomenon might explain the genetic structure pointed out with PCA.

Overall, the use of SE-GP and ME-GP models trained with MAGIC populations might find effective applications when the diversity of BPs originates from the same parents included in

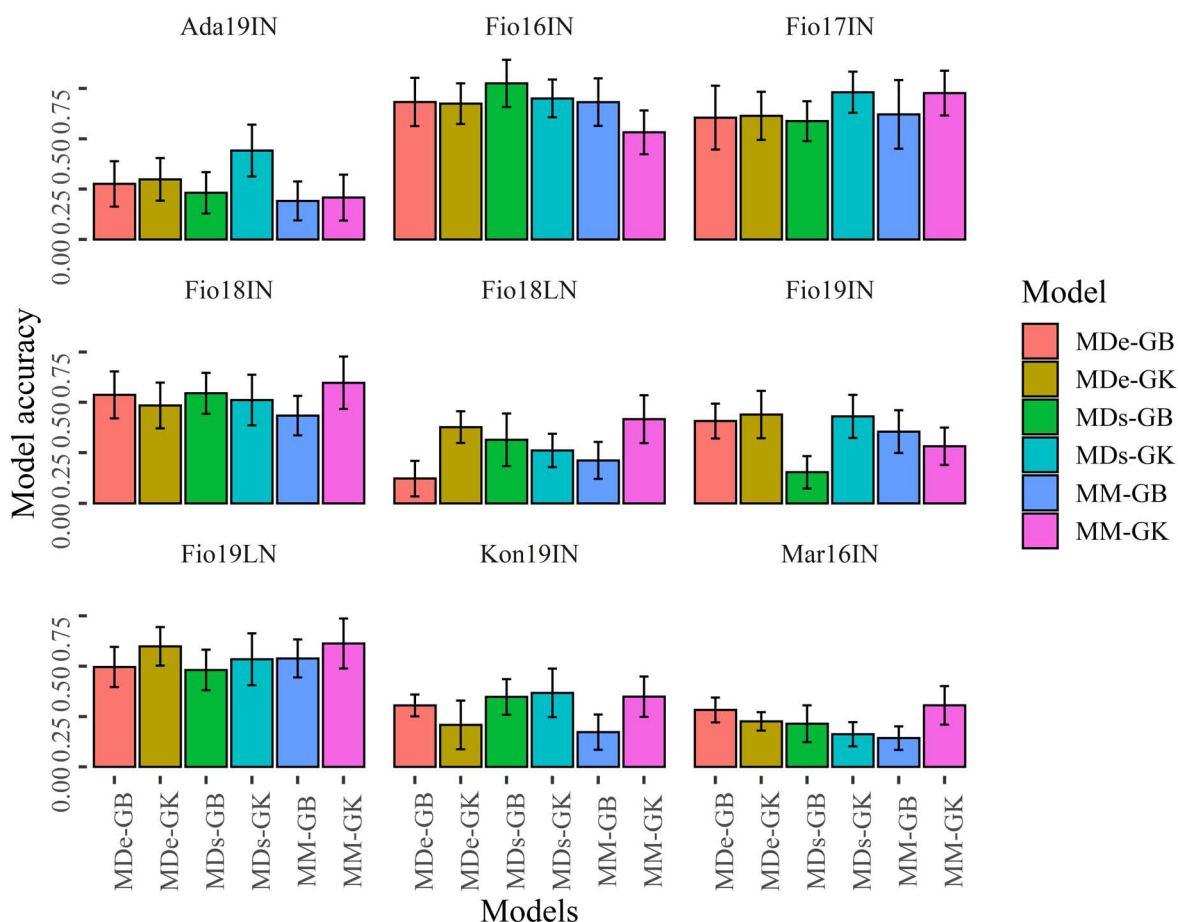


FIGURE 7 | Bar plots of the predictive ability values obtained with CV1. Bar plots show the mean correlation between observed and predicted values of GY obtained with 100 random CV1 partitions for MM, MDs, and MDe models implemented with GBLUP (GB) and Gaussian Kernel (GK) methods. Error bars point out the standard deviation of predictive ability values.

the founder set. In this case, GP models based on MAGIC populations might be applied to select the best offspring from crosses obtained with the MAGIC founders.

Benchmarking of Different TPs to Improve the Predictive Ability of GP Models

The composition of TPs and their genetic relationship with BPs affect the predictive ability of GP models as pointed out in several studies (Desta and Ortiz, 2014; Norman et al., 2018; Edwards et al., 2019) and to date several algorithms for optimizing TPs have been developed to increase the predictive ability of GP models (Akdemir et al., 2015). Untargeted and targeted optimization criteria based on GBLUP have been so far developed and tested in biparental populations and panel of mostly unrelated accessions. Nevertheless, the use of these optimization methods in actual breeding programs is hampered as the optimization process can lead to different optimized TP per each trait of interest. These optimization algorithms require *a priori* information (knowledge of the BP genotypes and traits

for which GP models must be developed) and output trait-dependent TPs (Akdemir et al., 2015). Moreover, in real breeding programs, BPs change over time and it might be difficult to implement these optimization procedures. Previous studies have shown that the relatedness between TPs and BPs has a large impact on the predictive ability of GP models, which can be improved increasing the genetic diversity of TPs (Norman et al., 2018). In fact, when the TP exhibits a narrow genetic diversity, low values of the predictive ability are often obtained in GP as it becomes impossible to predict all the marker effects that contribute to determining the phenotypic variations (Norman et al., 2018). Following these empirical findings, in this study we assembled a TP of 90 barley genotypes, which was named “TP-Diverse,” maximizing the genetic diversity among MAGIC lines and assessing its predictive ability using random CV schemes. Surprisingly, the predictive ability obtained with TP-Diverse was comparable with the predictive ability of GP models trained with the other three optimized TPs used in this study (Figure 2). One of the main advantages of using this approach is that the criterion adopted to assemble “TP-Diverse” depends only on genetic data and does not generate trait-dependent TPs. On the

other side, in this study we have not developed mathematical models to demonstrate or justify the rationale of this empirical criterion and consequently its validity should be further validated in other studies.

Fitting SE-GP and ME-GP Models Using the MAGIC Population of Barley

Several empirical analyses have been conducted to benchmark the predictive ability of different GP models in barley, maize and wheat panels of mostly unrelated accessions, biparental populations of *A. thaliana* and diallel crosses of maize and wheat to predict GY and other traits (Heslot et al., 2012). In this study, we presented another empirical analysis to assess the most promising GP models for MAGIC populations, implementing CV schemes for estimating the standard deviation of predictive ability values.

Three out five models fitted in this study (BayesA, BayesB, and BL) belong to the group of so called “Bayesian alphabet,” which denotes Bayesian linear regressions that differ in their prior density distribution (Gianola, 2013). In these Bayesian regression models, the prior density distribution assigned to marker effects controls the shrinkage of estimates and then different priors induce different types of shrinkage of marker effects. In the original description both BayesA and BayesB were introduced as hierarchical structures (Meuwissen et al., 2001) and it was later demonstrated that BayesA adopts a scaled t-distribution prior, while BayesB adopts priors that are mixtures of a peak in the vicinity of zero and of a continuous density priors (e.g., t, or normal density distribution) (Gianola et al., 2009). BL adopts a double exponential prior density distribution, which behaves similar to that of BayesA as both priors used in these models do not allow marker effects to be equal to zero and shrink estimates of the remaining marker effects. While the priors adopted in BL and BayesA prevent to have marker effects equal to zero, the prior used in BayesB allows to have null marker effects. The rationale of this prior is that in GP many markers might have a null contribution to the observed phenotypic variation. Although marker effects might be estimated differently, the predictive ability of the Bayesian models fitted in this study does not differ significantly (**Figure 1**). Moreover, our empirical analysis shows that the predictive ability of Bayesian models fitted to MAGIC populations is comparable with that of GB and GK models (**Figure 1**). Several empirical analyses have been carried out in cereal crops to highlight advantages and limits of different whole genome regression methods. In rice, SE-GP models fitted with BayesA, GB, and GK for three traits were compared using a reference panel of 284 accessions under different linkage disequilibrium scenarios (Ben Hassen et al., 2018). These results showed that under high linkage disequilibrium scenarios GK models slightly outperform GB in terms of prediction ability. Differently, when a subset of rice reference panel was used to predict the performance of 97 advanced lined derived from biparental crosses, GK and GB prediction ability showed comparable results for the three traits considered (Ben Hassen et al., 2018). Anyway, the results obtained in this study are limited to one (complex) trait and it

might plausible that for simpler traits GP models fitted in MAGIC might have different trend of the predictive ability.

Beyond SE-GP models, in this study we used the MAGIC population of barley to fit three different ME-GP models, two of which (MDs and MDe models) include terms for incorporating GxE interaction. In plant breeding, multi-environment field trials are routinely carried out to evaluate and exploit GxE interaction as it contributes to creating high-yielding genotypes. Consequently, modeling GxE interaction in GP has the potential to differentiate marker effects. MDe models used in this study (López-Cruz et al., 2015; Bandeira e Sousa et al., 2017) partition marker effects in main effects, that is effects that are stable across environments and environment-specific effects, that is interaction effects between markers and specific genotypes. As pointed out in other studies, MDe models are known to be more efficient when used along with sets of environments that have positive correlations. This limit arises as the pairwise correlation between environments is represented by the variance of the main marker effects, which in turn forces the co-variance between a pair of environments to be positive (López-Cruz et al., 2015; Bandeira e Sousa et al., 2017). This requirement is not trivial and might not allow to fit correctly MDe models. In our study, the adjusted means of GY in Mar16IN showed low or negative correlation with the other site-by-season combinations tested in this study and this might be the reason for which we have found that MDs models fit better the data, particularly when used in combination with the non-linear GK.

GP models based on reproducing kernel Hilbert Space along with the non-linear GK have the potential to capture non-additive genetic effects and potentially might outperform GB in terms of model fitting and predictive ability. In maize and wheat, comparison between the same GP models fitted with GB and the nonlinear GK for GY, unveiled that the latter method outperforms GB in terms of predictive ability in both single environment and multi-environment models (Cuevas et al., 2016; Bandeira e Sousa et al., 2017). In cereal crops, GY is a complex trait controlled by nonlinearity effects between genotypes and phenotypes owing to epistasis, environmental interactions (Bandeira e Sousa et al., 2017; Cuevas et al., 2018) and other interactions that are not considered in standard quantitative genetic models (Gianola et al., 2006). GK models have the potential to capture small and complex interactions, which are more evident in quantitative traits and this can explain the higher prediction ability of GK for GY. The empirical analysis presented in this study using barley MAGIC population corroborates that, for complex traits like GY, the predictive ability of GK outperforms that of GB. Overall, considering the number of models and methods fitted and the extensive field trials carried out across the Mediterranean, this study has delivered the most comprehensive empirical analysis of GP models fitted with MAGIC populations.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories

and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

AF designed and supervised the research along with the help of EI and LC. AF wrote the manuscript along with DP and significant contributions from AV, EI, HO, AC, AT, GV, LC, and AP. DP performed the research, while GV, AT, and SD developed the MAGIC population. AV, HO, EI, AC, SD, AT, IK, AP, and DP carried out field trials and plant phenotyping. All authors contributed to the article and approved the submitted version.

FUNDING

This research was carried out in the framework of the iBarMed project, which has been funded through the ARIMNet2 initiative and the Italian “Ministry of Agricultural, Food and Forestry Policies” under grant agreement “DM n. 20120.” ARIMNet2 has received funding from the EU 7th Framework Programme for research, technological development and demonstration under grant agreement no. 618127. The work was also supported by

SYSTEMIC_1063 (An integrated approach to the challenge of sustainable food systems: adaptive and mitigatory strategies to address climate change and malnutrition: From cereal diversity to plant breeding), a research project funded by Italian “Ministry of Agricultural, Food and Forestry Policies” in the frame of the Knowledge Hub on Food and Nutrition Security.

ACKNOWLEDGMENTS

This article is dedicated to the memory of Antonio Michele Stanca, who passed away on March 19th, 2020. Beyond supporting this study, Stanca’s research encompassed a wide range of plant science topics and contributed to exploiting morphological barley mutants to unveil the mechanisms that underlie plant developmental biology.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.664148/full#supplementary-material>

REFERENCES

- Afsharyan, N. P., Sannemann, W., Léon, J., and Ballvora, A. (2020). Effect of epistasis and environment on flowering time in barley reveals a novel flowering-delaying QTL allele. *J. Exp. Bot.* 71, 893–906. doi: 10.1093/jxb/erz477
- Akdemir, D., Sanchez, J. I., and Jannink, J. L. (2015). Optimization of genomic selection training populations with a genetic algorithm. *G. Select. Evol.* 47, 1–10. doi: 10.1186/s12711-015-0116-6
- Bandeira e Sousa, M., Cuevas, J., Couto, E. G., de, O., Pérez-Rodríguez, P., Jarquín, D., et al. (2017). Genomic-enabled prediction in maize using kernel models with genotype $\times$ environment interaction. *G3 Genes Genom. Genet.* 7, 1995–2014. doi: 10.1534/g3.117.042341
- Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., et al. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* 6:11. doi: 10.1186/1939-8433-6-11
- Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., and Crossa, J. (2015). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Sci.* 242, 23–36. doi: 10.1016/j.plantsci.2015.08.021
- Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bayer, M. M., Rapazote-Flores, P., Ganai, M., Hedley, P. E., Macaulay, M., Plieske, J., et al. (2017). Development and evaluation of a barley 50k iSelect SNP array. *Front. Plant Sci.* 8:1792. doi: 10.3389/fpls.2017.01792
- Ben Hassen, M., Cao, T. V., Bartholomé, J., Orasen, G., Colombi, C., Rakotomalala, J., et al. (2018). Rice diversity panel provides accurate genomic predictions for complex traits in the progenies of biparental crosses involving members of the panel. *Theoret. Appl. Genet.* 131, 417–435. doi: 10.1007/s00122-017-3011-4
- Berro, I., Lado, B., Nalin, R. S., Quincke, M., and Gutiérrez, L. (2019). Training population optimization for genomic selection. *Plant Genome* 12:190028. doi: 10.3835/plantgenome2019.04.0028
- Browning, B. L., and Browning, S. R. (2016). Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* 98, 116–126. doi: 10.1016/j.ajhg.2015.11.020
- Burguño, J., de los Campos, G., Weigel, K., and Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype $\times$ environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52, 707–719. doi: 10.2135/cropsci2011.06.0299
- Cavanagh, C., Morell, M., Mackay, I., and Powell, W. (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol.* 11, 215–221. doi: 10.1016/j.pbi.2008.01.002
- Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4:7. doi: 10.1186/s13742-015-0047-8
- Crossa, J., Pérez-rodríguez, P., Cuevas, J., Montesinos-lópez, O., Jarquín, D., Campos, G. D. L., et al. (2017). Genomic selection in plant breeding: methods. *Models Perspect.* 22, 961–975. doi: 10.1016/j.tplants.2017.08.011
- Cuevas, J., Crossa, J., Soberanis, V., Pérez-Elizalde, S., Pérez-Rodríguez, P., Campos, G., et al. (2016). Genomic prediction of genotype $\times$ environment interaction kernel regression models. *Plant Genome* 9:lantgenome2016.03.0024. doi: 10.3835/plantgenome2016.03.0024
- Cuevas, J., Granato, I., Fritsche-Neto, R., Montesinos-Lopez, O. A., Burguño, J., Sousa, M. B., et al. (2018). Genomic-enabled prediction Kernel models with random intercepts for multi-environment trials. *G3 Genes Genomes Genet.* 8, 1347–1365. doi: 10.1534/g3.117.300454
- De Beukelaer, H., Davenport, G. F., and Fack, V. (2018). Core hunter 3: flexible core subset selection. *BMC Bioinform.* 19:203. doi: 10.1186/s12859-018-2209-z
- De Los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., et al. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182, 375–385. doi: 10.1534/genetics.109.101501
- Dell’Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., et al. (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. *Genome Biol.* 16, 1–23. doi: 10.1186/s13059-015-0716-z
- Desta, Z. A., and Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* 19, 592–601. doi: 10.1016/j.tplants.2014.05.006
- Edwards, S. M. K., Buntjer, J. B., Jackson, R., Bentley, A. R., Lage, J., Byrne, E., et al. (2019). The effects of training population design on genomic prediction accuracy in wheat. *Theor. Appl. Genet.* 132, 1943–1952. doi: 10.1007/s00122-019-03327-y
- Emrich, K., Wilde, F., Miedaner, T., and Piepho, H. P. (2008). REML approach for adjusting the Fusarium head blight rating to a phenological date in inoculated

- selection experiments of wheat. *Theoret. Appl. Genet.* 117, 65–73. doi: 10.1007/s00122-008-0753-z
- Gianola, D. (2013). Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics* 194, 573–596. doi: 10.1534/genetics.113.151753
- Gianola, D., and Schon, C. C. (2016). Cross-validation without doing cross-validation in genome-enabled prediction. *G3 Genes Genom Genet.* 6, 3107–3128. doi: 10.1534/g3.116.033381
- Gianola, D., and Van Kaam, J. B. C. H. M. (2008). Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* 178, 2289–2303. doi: 10.1534/genetics.107.084285
- Gianola, D., De Los Campos, G., Hill, W. G., Manfredi, E., and Fernando, R. (2009). Additive genetic variability and the Bayesian alphabet. *Genetics* 183, 347–363. doi: 10.1534/genetics.109.103952
- Gianola, D., Fernando, R. L., and Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173, 1761–1776. doi: 10.1534/genetics.105.049510
- Gota, M., and Gianola, D. (2014). Kernel-based whole-genome prediction of complex traits: a review. *Front. Genet.* 5:363. doi: 10.3389/fgene.2014.00363
- Heffner, E. L., Jannink, J.-L., and Sorrells, M. E. (2011). Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4, 65–75. doi: 10.3835/plantgenome2010.12.0029
- Heffner, E. L., Sorrells, M. E., and Jannink, J. L. (2009). Genomic selection for crop improvement. *Crop Sci.* 49, 1–12. doi: 10.2135/cropsci2008.08.0512
- Heslot, N., Yang, H. P., Sorrells, M. E., and Jannink, J. L. (2012). Genomic selection in plant breeding: A comparison of models. *Crop Sci.* 52, 146–160. doi: 10.2135/cropsci2011.06.0297
- Hill, W. G., and Robertson, A. (2008). The effect of linkage on limits to artificial selection. *Genet. Res.* 89, 311–336. doi: 10.1017/S001667230800949X
- Huang, B. E., Verbyla, K. L., Verbyla, A. P., Raghavan, C., Singh, V. K., Gaur, P., et al. (2015). MAGIC populations in crops: current status and future prospects. *Theoret. Appl. Genet.* 128, 999–1017. doi: 10.1007/s00122-015-2506-0
- Jannink, J. L., Lorenz, A. J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genom. Proteom.* 9, 166–177. doi: 10.1093/bfpg/eqk001
- Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., et al. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoret. Appl. Genet.* 127, 595–607. doi: 10.1007/s00122-013-2243-1
- Jiménez-Galindo, J. C., Malvar, R. A., Butrón, A., Santiago, R., Samayoa, L. F., Caicedo, M., et al. (2019). Mapping of resistance to corn borers in a MAGIC population of maize. *BMC Plant Biol.* 19:431. doi: 10.1186/s12870-019-2052-z
- Kover, P. X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I. M., Purugganan, M. D., et al. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* 5:e1000551 doi: 10.1371/journal.pgen.1000551
- Laloe, D. (1993). Precision and information in linear models of genetic evaluation. *Genet. Select. Evol.* 25, 557–576. doi: 10.1051/gse:19930604
- López-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jannink, J.-L., et al. (2015). Increased prediction accuracy in wheat breeding trials using a marker $\times$ environment interaction genomic selection model. *G3 Genes Genom Genet.* 5, 569–582. doi: 10.1534/g3.114.016097
- Lorenzana, R. E., and Bernardo, R. (2009). Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoret. Appl. Genet.* 120, 151–161. doi: 10.1007/s00122-009-1166-3
- Mackay, I. J., Bansept-Basler, P., Bentley, A. R., Cockram, J., Gosman, N., Greenland, A. J., et al. (2014). An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. *G3 Genes Genom Genet.* 9, 1603–1610. doi: 10.1534/g3.114.012963
- Mathew, B., Léon, J., Sannemann, W., and Sillanpää, M. J. (2018). Detection of epistasis for flowering time using bayesian multilocus estimation in a barley MAGIC population. *Genetics* 208, 525–536. doi: 10.1534/genetics.117.300546
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Monat, C., Padmarasu, S., Lux, T., Wicker, T., Gundlach, H., Himmelbach, A., et al. (2019). TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools. *Genome Biol.* 20:284.
- Norman, A., Taylor, J., Edwards, J., and Kuchel, H. (2018). Optimising genomic selection in wheat: effect of marker density, population size and population structure on prediction accuracy. *G3 Genes Genom Genet.* 8, 2889–2899. doi: 10.1534/g3.118.200311
- Ou, J. H., and Liao, C. T. (2019). Training set determination for genomic selection. *Theoret. Appl. Genet.* 132, 2781–2792. doi: 10.1007/s00122-019-03387-0
- Park, T., and Casella, G. (2008). The bayesian lasso. *J. Am. Statist. Assoc.* 103, 681–686. doi: 10.1198/016214508000000337
- Pérez, P., and de los Campos, G. (2014). BGLR: a statistical package for whole genome regression and prediction. *Genetics* 198, 483–495. doi: 10.1534/genetics.114.164442
- Pérez-Elizalde, S., Cuevas, J., Pérez-Rodríguez, P., and Crossa, J. (2015). Selection of the bandwidth parameter in a bayesian kernel regression model for genomic-enabled prediction. *J. Agricult. Biol. Environ. Statist.* 20, 512–532. doi: 10.1007/s13253-015-0229-y
- Ponce, K. S., Ye, G., and Zhao, X. (2018). QTL identification for cooking and eating quality in indica rice using multi-parent advanced generation intercross (MAGIC) population. *Front. Plant Sci.* 9:868. doi: 10.3389/fpls.2018.00868
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- R Core Team (2019). *R: A Language and Environment for Statistical Computing. Industrial and Commercial Training*. Vienna: R Core Team. doi: 10.1108/eb003648
- Rebetzke, G. J., Verbyla, A. P., Verbyla, K. L., Morell, M. K., and Cavanagh, C. R. (2014). Use of a large multiparent wheat mapping population in genomic dissection of coleoptile and seedling growth. *Plant Biotechnol. J.* 12, 219–230. doi: 10.1111/pbi.12130
- Riaz, A., KockAppelgren, P., Hehir, J. G., Kang, J., Meade, F., Cockram, J., et al. (2020). Genetic analysis using a multi-parent wheat population identifies novel sources of septoria tritici blotch resistance. *Genes* 11, 1–26. doi: 10.3390/genes11080887
- Rincint, R., Laloë, D., Nicolas, S., Altmann, T., Brunel, D., Revilla, P., et al. (2012). Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics* 192, 715–728. doi: 10.1534/genetics.112.141473
- Sannemann, W., Lisker, A., Maurer, A., Léon, J., Kazman, E., Cöster, H., et al. (2018). Adaptive selection of founder segments and epistatic control of plant height in the MAGIC winter wheat population WM-800. *BMC Genomics* 19:3. doi: 10.1186/s12864-018-4915-3
- Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, S. A., et al. (2020). Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity* 125, 396–416. doi: 10.1038/s41437-020-0336-6
- Stadlmeier, M., Hartl, L., and Mohler, V. (2018). Usefulness of a multiparent advanced generation intercross population with a greatly reduced mating design for genetic studies in winter wheat. *Front. Plant Sci.* 871:e01825. doi: 10.3389/fpls.2018.01825
- Thachuk, C., Crossa, J., Franco, J., Dreisigacker, S., Warburton, M., and Davenport, G. F. (2009). Core hunter: an algorithm for sampling genetic resources based on multiple genetic measures. *BMC Bioinformatics* 10:243. doi: 10.1186/1471-2105-10-243
- Thioulouse, J., Dufour, A. B., Jombart, T., Dray, S., Siberchicot, A., and Pavoine, S. (2018). Multivariate analysis of ecological data with ade4. *Multivar. Anal. Ecol. Data ade4* 18:58850. doi: 10.1007/978-1-4939-8850-1
- Tishbirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. R. Statist. Soc. Ser. B (Methodol.)* 58, 267–288.
- Valdar, W., Flint, J., and Mott, R. (2006). Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics* 172, 1783–1797. doi: 10.1534/genetics.104.039313
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Wang, X., Xu, Y., Hu, Z., and Xu, C. (2018). Genomic selection methods for crop improvement: current status and prospects. *Crop J.* 6, 330–340. doi: 10.1016/j.cj.2018.03.001
- Watson, A., Ghosh, S., Williams, M. J., Cuddy, W. S., Simmonds, J., Rey, M. D., et al. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat. Plants* 4, 23–29. doi: 10.1038/s41477-017-0083-8

- Wickham, H. (2016). ggplot 2: Elegant graphics for data analysis. *Media* 35:98141. doi: 10.1007/978-0-387-98141-3
- Wimmer, V., Albrecht, T., Auinger, H. J., and Schön, C. C. (2012). Synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* 28, 2086–2087. doi: 10.1093/bioinformatics/bts335
- Xu, X., Sharma, R., Tondelli, A., Russell, J., Comadran, J., Schnaithmann, F., et al. (2018). Genome-wide association analysis of grain yield-associated traits in a pan-european barley cultivar collection. *Plant Genome* 11:170073. doi: 10.3835/plantgenome2017.08.0073
- Yalcin, B., Flint, J., and Mott, R. (2005). Using progenitor strain information to identify quantitative trait nucleotides in outbred mice. *Genetics* 171:673–81. doi: 10.1534/genetics.104.028902
- Zadoks, J. C., Chang, T. T., and Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Res.* 14, 415–421. doi: 10.1111/j.1365-3180.1974.tb01084.x
- Zaw, H., Raghavan, C., Pocsedio, A., Swamy, B. P. M., Jubay, M. L., Singh, R. K., et al. (2019). Exploring genetic architecture of grain yield and quality traits in a 16-way indica by japonica rice MAGIC global population. *Sci. Rep.* 9, 1–11. doi: 10.1038/s41598-019-55357-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Puglisi, Delbono, Visioni, Ozkan, Kara, Casas, Igartua, Valè, Piero, Cattivelli, Tondelli and Fricano. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Identification, Characterization, and Expression Profile Analysis of the *mTERF* Gene Family and Its Role in the Response to Abiotic Stress in Barley (*Hordeum vulgare* L.)

Tingting Li^{1†}, Wenqiu Pan^{2†}, Yiyuan Yuan¹, Ying Liu¹, Yihan Li¹, Xiaoyu Wu¹, Fei Wang¹ and Licao Cui^{1*}

¹ College of Bioscience and Engineering, Jiangxi Agricultural University, Nanchang, China, ² State Key Laboratory of Crop Stress Biology in Arid Areas and College of Agronomy, Northwest A&F University, Yangling, China

OPEN ACCESS

Edited by:

Hakan Ozkan,
Çukurova University, Turkey

Reviewed by:

Huseyin Tombuloglu,
Imam Abdulrahman Bin Faisal
University, Saudi Arabia
Victor Quesada,
Miguel Hernández University of
Elche, Spain

*Correspondence:

Licao Cui
cuilicao@jxau.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 09 April 2021

Accepted: 23 June 2021

Published: 15 July 2021

Citation:

Li T, Pan W, Yuan Y, Liu Y, Li Y, Wu X,
Wang F and Cui L (2021)
Identification, Characterization, and
Expression Profile Analysis of the
mTERF Gene Family and Its Role in
the Response to Abiotic Stress in
Barley (*Hordeum vulgare* L.).
Front. Plant Sci. 12:684619.
doi: 10.3389/fpls.2021.684619

Plant mitochondrial transcription termination factor (*mTERF*) family regulates organellar gene expression (OGE) and is functionally characterized in diverse species. However, limited data are available about its functions in the agriculturally important cereal barley (*Hordeum vulgare* L.). In this study, we identified 60 *mTERFs* in the barley genome (*HvmTERFs*) through a comprehensive search against the most updated barley reference genome, Morex V2. Then, phylogenetic analysis categorized these genes into nine subfamilies, with approximately half of the *HvmTERFs* belonging to subfamily IX. Members within the same subfamily generally possessed conserved motif composition and exon-intron structure. Both segmental and tandem duplication contributed to the expansion of *HvmTERFs*, and the duplicated gene pairs were subjected to strong purifying selection. Expression analysis suggested that many *HvmTERFs* may play important roles in barley development (e.g., seedlings, leaves, and developing inflorescences) and abiotic stresses (e.g., cold, salt, and metal ion), and *HvmTERF21* and *HvmTERF23* were significant induced by various abiotic stresses and/or phytohormone treatment. Finally, the nucleotide diversity was decreased by only 4.5% for *HvmTERFs* during the process of barley domestication. Collectively, this is the first report to characterize *HvmTERFs*, which will not only provide important insights into further evolutionary studies but also contribute to a better understanding of the potential functions of *HvmTERFs* and ultimately will be useful in future gene functional studies.

Keywords: barley, *mTERF* gene family, duplication, expression profile, qRT-PCR, genetic variation

INTRODUCTION

One of the major differences between eukaryotes and prokaryotes is that the former has organelles, while the latter does not (Quesada, 2016). Due to endosymbiotic evolution from their cyanobacterial ancestors, most of the organellar genes within chloroplasts and mitochondria have been either lost or transferred to the nucleus (Gray, 2012). Current chloroplast and mitochondrial genomes retain only a tiny fraction of the genes, which are required for photosynthesis, gene

expression, and electron transport chains (Lang et al., 1999; Yagi and Shiina, 2012; Robles and Quesada, 2021). Nevertheless, thousands of proteins have been predicted to be localized in plant mitochondria and chloroplasts according to bioinformatics analysis, most of which are encoded by the nuclear genome (Binder and Brennicke, 2003; Huang et al., 2013; Lee et al., 2013). Thus, the organellar gene expression (OGE) apparatus is a precisely coordinated system that largely depends on a great many of proteins encoded by nuclear genes (Pfannschmidt et al., 2015; Quesada, 2016). In the mitochondria and chloroplasts of higher plants, several transcriptional components of the functional OGE system have been reported. Three different polymerases involved in the transcriptional machinery have been demonstrated, including a multi-subunit plastid-encoded RNA polymerase (PEP) and two single-subunit nucleus-encoded RNA polymerases (NEPs) (Pfannschmidt et al., 2015). However, currently known auxiliary factors can only partially explain the transcriptional machinery, suggesting the existence of additional unidentified regulatory factors that are required for organellar gene transcription (Kühn et al., 2007; Liere et al., 2011).

Among the nucleus-encoded OGE factors, a novel protein family has received increasing concerns, namely, the mitochondrial transcription termination factor (mTERF) family. mTERF proteins, firstly characterized in animal mitochondria, were involved in mitochondrial transcription, translation, and DNA replication (Quesada, 2016). These proteins possess a variable number of ~30 amino acid “mTERF” motifs and comprise three leucine zipper-like elements separated by loops (Roberti et al., 2006), which are believed to confer the ability to recognize and bind to the typical mTERF motif on mitochondrial genome (Roberti et al., 2009). To date, four members, mTERF1 to mTERF4, have been described in vertebrates (Linder et al., 2005; Roberti et al., 2009). mTERF1, the founding member of this family, was originally considered to promote transcription termination of the heavy strand genes *tRNA-Ler* and *16S rRNA* (Kruse et al., 1989). However, more recent studies proposed that mTERF1 only partially terminated heavy strand

transcription, whereas its major function was to completely block transcriptional interference at the opposite light strand of the ribosomal RNA genes from which they originated (Terzioglu et al., 2013). Although mTERF2 is a non-specific mitochondrial DNA binding protein and works as a negative regulator of mitochondrial gene expression, the function of mTERF2 is largely unknown (Pellegrini et al., 2009; Huang et al., 2011). Similarly, mTERF3 has been demonstrated as a specific repressor of mammalian mitochondrial transcription initiation, and therefore slowing down cell metabolism (Park et al., 2007). Meanwhile, other studies also revealed that mTERF3 was essential for ribosome biogenesis, mitochondrial protein transcription, and translation (Andersson et al., 2011; Wredenberg et al., 2013). mTERF4 can directly regulate mitochondrial ribosomal biogenesis and protein translation by targeting to the ribosomal RNA methyltransferase NSUN4 (a 5-methylcytosine RNA methyltransferase) (Cámara et al., 2011; Spähr et al., 2012; Yakubovskaya et al., 2012).

By contrast, the *mTERF* gene family has expanded to approximately 30 members during the evolutionary process of land plants (Quesada, 2016; Leister and Kleine, 2020). For example, there are 35 *mTERFs* in *Arabidopsis thaliana*, 33 in rice (*Oryza sativa*) (Kleine, 2012), 31 in maize (*Zea mays*) (Zhao et al., 2014), 25 in grape (*Vitis vinifera*) (Yin et al., 2021), and 35 in *Capsicum annuum* (Tang et al., 2019). The substantial expansion in the number of *mTERF* genes was accompanied by their increased involvement in diverse RNA metabolism processes, with the majority being involved in rRNA maturation and intron splicing in organelles (Méteignier et al., 2020). For instance, *Arabidopsis mTERF8* mediates preferential transcription termination of the chloroplast gene *psbJ* by preferentially binding to the 3'-terminus (Xiong et al., 2020). *Arabidopsis mTERF15* acts as an RNA binding protein that is required for mitochondrial *nad2* intron 3 splicing and functional complex I activity, which is indispensable for plant growth and development (Hsu et al., 2014). mTERF6 is required for the maturation of the chloroplast Ile transfer RNA gene *trnI.2* and regulates transcription termination of the PEP core subunit *rpoA* poly-cistron, thus further demonstrating the essential roles of mTERFs in leaf organogenesis and patterning in *Arabidopsis* (Romani et al., 2015; Robles et al., 2018b; Zhang et al., 2018). A more recent study reported that mTERF2 is implicated in the splicing of the group IIB introns of *ycf3* (intron 1) and *rps12* in *Arabidopsis*. Knock-down *mTERF2* resulted in delayed flowering time and knock-out *mTERF2* mutants were embryo lethal (Lee et al., 2021). In maize, ZmmTERF4 is involved in plastid ribosome accumulation and promote group II intron splicing of *trnI.2*, *trnA*, *rpl2*, *atpE*, and *ycf3-2* in chloroplasts (Hammani and Barkan, 2014). Zmsmk3 affects complex I assembly by modulating *nad4* intron 1 and *nad1* intron 4 splicing, seedling growth, and kernel development (Pan et al., 2019). Moreover, recent studies have also proposed the importance of *mTERF* genes associated with a variety of abiotic stress responses, including heat, salt, and osmotic stresses. For instance, *Arabidopsis* SHOT1 (SUPPRESSOR OF HOT1-4 1) can indirectly increase thermotolerance by reducing reactive oxygen species (ROS) accumulation and increasing the expression of heat shock proteins (HSPs), particularly those localized to

Abbreviations: ABA, abscisic acid; ABRE, ABA-responsive elements; Aux/IAA, auxin/indole-3-acetic acid; BLAST, basic local alignment search tool; C3H, cysteine3histidine; CDD, conserved domains database; CDS, coding sequence; EST, expressed sequence tag; FDR, false discovery rate; FPKM, fragments per kilobase per million; ERE, estrogen response element; Fst, Wright's F-statistic; GCN4, general control non-repressible-4; GO, gene ontology; GRAVY, grand average of hydropathicity; GRF, growth-regulating factor; GSDS, Gene Structure Display Server; GTF, Gene Transfer Format; HMM, hidden Markov model; HSP, heat shock protein; Ka, non-synonymous substitution rate; Ks, synonymous substitution rate; LTP, lipid transfer protein; LTR, long terminal repeat; MBS, myeloblastosis binding site; MDA1, mTERF DEFECTIVE IN *Arabidopsis*1; MeJA, methyl jasmonate; MEME, Multiple Em for Motif Elicitation; MSA-like, mitosis-specific activator; MW, molecular weight; MYB, myeloblastosis; NCBI, National Coalition Building Institute; NEP, nucleus-encoded RNA polymerase; NJ, neighbor-joining; OGE, organellar gene expression; PAML, Phylogenetic Analysis by Maximum Likelihood; PCA, principal component analysis; PEP, plastid-encoded RNA polymerase; pI, theoretical isoelectric point; qRT-PCR, Quantitative Real-time PCR; ROS, reactive oxygen species; SA, salicylic acid; SHOT1, SUPPRESSOR OF HOT1-4 1; SL1, SEEDLING LETHAL 1; SMART, simple modular architecture research tool; SNP, single nucleotide polymorphism; SOLDAT10, SINGLET OXYGEN-LINKED DEATH ACTIVATOR10; SRA, sequence read archive; Ts, transition; Tv, transversion; WGCNA, weighted gene co-expression network analysis.

mitochondria (Kim et al., 2012). *mTERF5/MDA1* (*mTERF* DEFECTIVE IN *Arabidopsis1*) not only has a dual function in the transcription and stabilization of specific chloroplast transcripts but also responds to salt, osmotic, and sugar stresses through perturbed abscisic acid (ABA) retrograde signaling during seedling establishment in *Arabidopsis* (Robles et al., 2012; Ding et al., 2019; Méteignier et al., 2020). *Arabidopsis mTERF9* regulates chloroplast gene expression and development and responds to sugar, ABA, salt, and osmotic stresses (Robles et al., 2015). Similar to *mTERF5* and *mTERF9*, loss of *Arabidopsis mTREF27* resulted in mitochondria developmental defects and altered response to salt stress (Jiang et al., 2021). *Arabidopsis mTERF10* and *mTERF11* are involved in the response to salt stress, possibly through the ABA-mediated pathway (Xu et al., 2017). Recently, a new role was demonstrated for *mTERF6* in response to adverse environmental stresses, such as ABA, ionic, and osmotic stresses (Robles et al., 2018a). *Arabidopsis SOLDAT10* (SINGLET OXYGEN-LINKED DEATH ACTIVATOR10) controls plastid-specific rRNA expression and protein synthesis in plastids and is well known for its roles in the response to mild photooxidative stress (Meskauskiene et al., 2009). Collectively, *mTERFs* are essential for the regulation of OGE and play crucial roles in plant growth and development and in response to diverse abiotic stresses, at least in *Arabidopsis* and possibly in other higher plants. Nevertheless, detailed information about the molecular mechanisms of *mTERFs* is still rather limited in diverse plants, especially crop plants (Zhao et al., 2014).

As one of the earliest domesticated crops of ancient civilizations, barley (*Hordeum vulgare* L.) currently ranks as the fourth most abundant crop in terms of both area and tonnage harvested (Mayer et al., 2012). Barley is more adaptable to a wide range of agroclimatic conditions than its relative wheat and, as a result, is of high importance for human food, animal feed, and malt brewing (Jayakodi et al., 2020). The first draft sequence assembly of barley (Mayer et al., 2012) and its subsequent improved versions (Mascher et al., 2017; Monat et al., 2019) lay the foundation for the comprehensive identification and characterization of gene families at the genome-wide level. Here, the protein sequences of barley *mTERFs* were identified through a comprehensive search. The physicochemical properties, phylogenetic relationships, exon-intron gene structure, conserved motifs, expression profiles, and preliminary functions were systematically analyzed. Moreover, the single-nucleotide polymorphism (SNP) variation atlas of *mTERFs* for wild and landrace barley accessions was profiled. This study will not only shed light on the evolutionary mechanism of barley *mTERFs*, but also pave the way for their functional characterization in barley and beyond.

MATERIALS AND METHODS

Identification of *mTERF* Gene Family Members in Barley

The protein sequences of barley Morex V2 were downloaded from the IPK database (<https://doi.org/10.5447/ipk/2019/8>),

and the hidden Markov model (HMM) file of the *mTERF* domain (PF02536) was retrieved from the Pfam database. HMMER v2.41.1 was employed to search for the *mTERF* domain against the barley genome with the default inclusion threshold. The candidate sequences were further confirmed by using the NCBI-CDD (National Coalition Building Institute, Conserved Domains Database) (<https://www.ncbi.nlm.nih.gov/cdd/>), SMART (Simple Modular Architecture Research Tool) (<http://smart.embl-heidelberg.de/>), HMMER (<https://www.ebi.ac.uk/Tools/hmmer/>), and InterPro (<http://www.ebi.ac.uk/interpro/search/sequence/>) online tools. Subsequently, a BLAST (Basic Local Alignment Search Tool) search against barley ESTs (Expressed Sequence Tag) was conducted to determine the existence of putative *mTERF* genes. The molecular weight (MW), number of amino acids, theoretical isoelectric point (pI), and grand average of hydropathicity (GRAVY) were evaluated using the online tool ExPASy (<http://web.expasy.org/protparam/>). The subcellular localization was predicted using the TargetP online tools (<http://www.cbs.dtu.dk/services/TargetP/>).

Phylogenetic Relationship, Gene Structure, and Conserved Motif Analysis

Multiple sequence alignment of full-length proteins of *HvmTERF* genes was performed using the Clustal X program. An unrooted neighbor-joining (NJ) phylogenetic tree was constructed using MEGA X with 1,000 bootstrap replicates. The exon-intron gene structure was visualized using the Gene Structure Display Server (GSDS) (<http://gsds.cbi.pku.edu.cn/>) based on the gene annotation GTF (Gene Transfer Format) file. The conserved protein motifs were obtained using online MEME (Multiple Em for Motif Elicitation) tools (<https://meme-suite.org/meme/>) with the following parameters: the maximum number of motifs was set to 10, any number of repetitions was allowed, and the optimum width ranged from 6 to 250. The 1.5 kb genomic sequences upstream of the coding regions were extracted and submitted to the PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) to identify the putative *cis*-acting regulatory elements in the promoter region. The transcripts of *HvmTERFs* were extracted and submitted to the psRNATarget online server (<http://plantgrn.noble.org/psRNATarget/>) to detect the candidate miRNA targets with the following parameters: published miRNAs from *Brachypodium*, barley, and wheat were chosen, with a maximum expectation = 4.

Gene Duplication and Comparative Genomics Analysis of Barley, *Arabidopsis*, *Brachypodium*, Rice, Grape, and Maize *mTERFs*

To reveal the duplication events of *HvmTERF* during barley evolution, an integrated method was employed to identify the duplicated pairs. First, MCScanX software was used to detect duplication events. Second, the following criteria were used as described by Chen et al. (1) the alignment of shorter genes covered $\geq 70\%$ of longer genes; (2) the aligned region possessed an identity $\geq 70\%$; and (3) only one duplication event was counted for tightly linked genes (Gu et al., 2002; Chen

et al., 2012). The duplicated events were manually combined into a non-redundant dataset to determine the orthologous relationships between barley and other species. The orthologs of *mTERF* genes in *A. thaliana*, *Brachypodium distachyon*, *O. sativa*, *V. vinifera*, and *Z. mays* were identified using InParanoid V4.1. The syntenic blocks within and among species were detected by MCscanX. To evaluate the evolutionary rate of the duplicated and syntenic genes, PAML (Phylogenetic Analysis by Maximum Likelihood) v4.3 software was utilized to calculate the non-synonymous (Ka) and synonymous (Ks) substitution ratios. The duplicated pairs were visualized using Circos v0.67 software.

Expression Analysis of *HvmTERF* Genes

To estimate the gene expression profile of *HvmTERFs*, RNA-seq samples from different tissues and developmental stages as well as plants responding to various biotic and abiotic stresses were retrieved from the NCBI Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/>). The sample information and accession numbers are shown in **Supplementary Table 1**. The Hisat2 v2.1.0 and StringTie v1.3.5 pipelines were employed to calculate the fragments per kilobase of transcript per million fragments mapped (FPKM) value. The R package Ballgown was used to identify the differentially expressed genes. The differentially expressed genes were identified as having a false discovery rate (FDR) ≤ 0.05 and fold change ≥ 2 . The heatmap and hierarchical clustering were generated using the pheatmap package embedded in R with the log2 transformed FPKM values. To determine the co-expressed genes with *HvmTERFs*, a co-expression network was constructed by weighted gene co-expression network analysis (WGCNA) in R. Here, a convenient one-step method was employed for network construction, and genes with the top 5% weighted values associated with *HvmTERFs* were categorized for further analysis. A BLAST search against *Arabidopsis* and rice proteins was performed to determine the potential functions of the co-expressed genes. Cytoscape v3.8.0 was implemented to display the co-expression networks.

Plant Materials, Treatment, and qRT-PCR Analysis

Seeds of the barley cultivar Morex were sterilized with 5% sodium hypochlorite for 10 min, rinsed with distilled water, and then germinated on wet filter paper at 25°C for 5 days. The germinated seeds were hydroponically cultured in a greenhouse under the following conditions: 20°C day/15°C night, 16 h light/8 h dark cycle, and 50% relative humidity. Three-leaf-stage seedlings were exposed to 150 mM NaCl, 20% PEG, 4°C, or 100 μ M ABA for 0, 1, 6, 12, 24, and 48 h. Seedlings without any treatment at the same time point were used as the control. Leaves and roots were collected from three plants at each time point and promptly frozen in liquid nitrogen for RNA extraction with three biological replicates.

To further investigate the possible functions of *HvmTERFs*, a total of 25 *HvmTERFs* were randomly selected to detect their expression patterns through qRT-PCR (Quantitative Real-time PCR) analysis. The primers used in this study are listed in **Supplementary Table 2**. Total RNA was isolated using a Plant

RNA extraction kit (Omega Biotek, USA), and cDNA was synthesized using 5X All-in-one RT MasterMix (ABM, Canada) following the manufacturer's instructions. *HvACTIN2* (GenBank accession no. AY145451.1) was used as the internal control. The TB-Green® Premix Ex Taq™ II kit (Takara, Dalian, China) was used for qRT-PCR amplification in a QuantStudio™ Real-Time PCR system (Thermo Fisher, USA). The reaction protocol was as follows: 95°C for 30 s, followed by 40 cycles at 95°C for 3 s and 60°C for 30 s. The relative expression levels of candidate genes were calculated using the $2^{-\Delta\Delta CT}$ method. Three technical replicates were applied for each treatment (Livak and Schmittgen, 2001). Student's *t*-test was employed for statistical analysis by R software. The histogram was drawn using the ggplot package in R software. One asterisk (*) and double asterisk (**) indicate 0.05 and 0.01 significance level, respectively.

Nucleotide Variation, Population Structure, and Haplotype Analysis of *HvmTERFs*

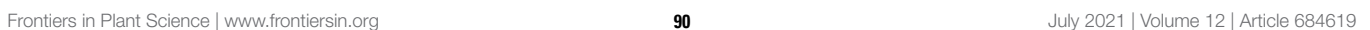
To acquire the candidate *HvmTERFs*, a total of 220 barley resequencing samples were downloaded from the SRA database (Russell et al., 2016). The geographic distribution is presented in **Supplementary Figure 1**. The detailed material information is listed in **Supplementary Table 3**. BWA-MEM v0.7.13r1126 was used to map the clean reads against the barley reference genome. The PICARD-GATK pipeline was employed to generate single nucleotide polymorphisms (SNPs) with default parameters. The genomic distribution and potential function of the SNPs were annotated by SnpEFF v4.3. The SNPs located within the *HvmTERF* genes were retained for subsequent analysis. To further reveal the population structure of barley samples based on *HvmTERF* sequences, population structure analysis, phylogenetic tree analysis, and principal component analysis (PCA) were performed. ADMIXTURE v1.3.0 was used to infer the population structure with predefined K-values ranging from 2 to 5. The phylogenetic tree was constructed using Treebest v1.9.2. The Smartpca toolkit implemented in EIGENSOFT v4.2 was employed to conduct the PCA. Median-joining haplotype networks were constructed using the software programs DnaSP v5.10.01, Alignment v1.3.1.1, and Network v4.6.1.1. The network was visualized using Cytoscape v3.8.0. The nucleotide diversity (π) and Wright's F-statistic (Fst) were calculated using vcftools v0.1.16.

RESULTS

Identification of *mTERF* Gene Family Members in Barley

The updated reference genome of barley, Morex v2, provided invaluable resources for *HvmTERF* identification, and a total of 60 *mTERF* genes were identified in barley using a combined method (**Supplementary Tables 4, 5**). Since there was no standard nomenclature, the barley *mTERFs* were designated as *HvmTERF1* to *HvmTERF60* according to their chromosome numbers and physical positions. The physicochemical properties of the *HvmTERFs* were further characterized. In detail, the *mTERFs* encoded proteins ranging from 105 (*HvmTERF41*) to

localized to mitochondria (39 HvmTERFs, 65%) or chloroplasts (8 HvmTERFs, 13.3%), and the remaining 13 HvmTERFs were targeted to other locations. To confirm the existence of *HvmTERFs*, a BLAST search against barley ESTs was performed. In total, 48 members of the *HvmTERF* gene family had EST records, whereas the remaining 12 *mTERFs* had no EST support, suggesting their stage- or tissue-specific expression profile or undetectable expression level.

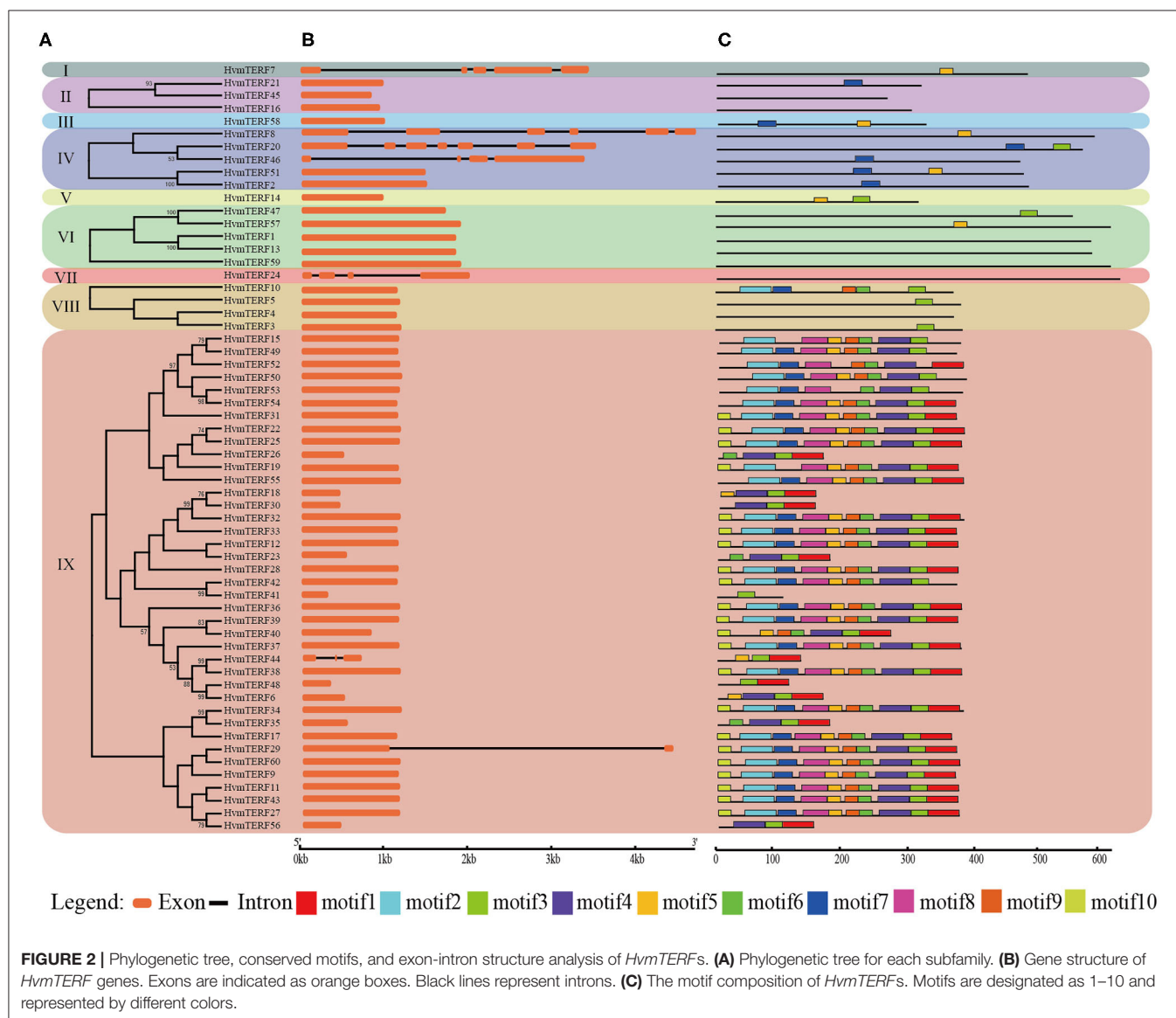


Phylogenetic and Structural Domain Analysis of HvmTERFs

We examined the amino acid sequence features of the mTERF domain by multiple sequence alignment. The conserved mTERF motifs spanned approximately 30 amino acids in length, have been characterized in other plants and are believed to act as DNA-binding modules (Zhao et al., 2014) (Supplementary Figure 2; Supplementary Table 6). The sequence conservation percentages for each amino acid residue were calculated, and 15 amino acid sites were highly conserved with a consensus sequence percentage >60%. Consistent with previous studies (Zhao et al., 2014), three repeats of the leucine zipper-like heptad X₃LX₃ were identified in barley mTERF motifs, of which the conservation percentages were 62.71, 46.67, and 36.67% for Leu-8, Leu-16, and Leu-23, respectively. These results revealed that HvmTERFs possessed well-characterized

mTERF motifs with conserved leucine residues like those in other plants, indicating the conserved evolutionary process of plant mTERF proteins. Surprisingly, the conservation percentages of Ile-1 and Tyr-20 in barley were significantly higher than those in *Arabidopsis*, rice, and maize, suggesting that these residues may play essential roles in the evolutionary history of *HvmTERFs*.

To further elucidate the evolutionary relationship of *HvmTERFs*, we constructed a phylogenetic tree based on the alignment of 128 mTERF protein sequences from *Arabidopsis* (35), rice (33), and barley (60) (Figure 1). These mTERF proteins were divided into nine monophyletic clades according to the classification given by Zhao (Zhao et al., 2014). The number of proteins assigned to different subfamilies varied greatly, of which subfamily IX contained 39 members, whereas subfamilies I, III, V, and VII possessed only one member.



The exon-intron structure not only provides additional evidence to support the phylogenetic topology but also increases the understanding of the functional diversification within a gene family. Therefore, the exon-intron structure of *HvmTERFs* was analyzed to obtain their evolutionary relationships (Figure 2). A solid correlation between gene structures and their phylogeny was observed. Genes clustered within the same subfamily displayed a similar exon-intron structure. Indeed, *HvmTERFs* within subfamilies II, VI, and VIII showed nearly identical exon lengths and tended to be intron-less. Nonetheless, we pinpointed the exon/intron gain/loss event within several clusters. For example, 37 out of 39 *HvmTERFs* within subfamily IX possessed only 1 exon, whereas *HvmTERF29* and *HvmTERF44* had 2 and 3 exons, indicating that they may have acquired additional exons during the evolutionary history of the *mTERF* gene family.

To further gain insight into the evolutionary relationships and functional regions among the *HvmTERFs*, the distribution pattern of the conserved motifs was also visualized (Figure 2). We identified 10 motifs and designated them as motif 1 to motif 10. Notably, no motif was identified within *HvmTERFs* 1, 4, 13, 16, 24, 45, and 59, possibly because the consensus sequence failed to reach the threshold in the MEME software. The *HvmTERFs* within the same clade showed similar motif numbers and distribution patterns. Motif 3 was shared by 44 members, ranking as the most abundant motif, followed by motif 4 (36) and motif 5 (36). Except for motifs 2, 5, 6, 7, and 9, the remaining motifs were specific to subfamily IX. We also observed a certain order of the identified motifs. For example, motif 2 tended to be tightly connect with motif 7, motif 5 was linked to motif 8, motif 6 was linked to motif 9, and motifs 1, 3, and 4 were linked.

Duplication Events and Orthologous Analysis of *HvmTERFs*

The *HvmTERFs* were unevenly located across the seven barley chromosomes in accordance with the barley genome annotation, of which 27 *HvmTERFs* were located on chromosome 6, ranking as the most populated chromosome, whereas the other six chromosomes had only 9 (chromosome 7H) to 3 (chromosome 3H) *HvmTERF* genes (Supplementary Figure 3). Interestingly, there was no positive correlation between chromosome length and the number of *HvmTERFs* (Pearson correlation $r = -0.2994$, p -value = 0.5141), indicating that longer chromosomes do not necessarily contain more *HvmTERF* genes.

In order to elucidate the expansion mechanism of *HvmTERFs*, tandem, and segmental duplication event analyses were performed using an integrated method (Figure 3). The results showed that 10 *HvmTERFs* (*HvmTERFs* 25 and 26, *HvmTERFs* 29 and 32, *HvmTERFs* 49 and 50, and *HvmTERFs* 36, 37, 38, and 39) were clustered into four tandemly duplicated regions on chromosome 6, indicating a gene distribution hot spot of *HvmTERFs*. It is noteworthy that all tandemly duplicated genes belonged to subfamily IX. Generally, it is difficult to segregate this kind of tightly linked gene arrangement through recombination in breeding or research. Furthermore, seven duplicated pairs composed of 13 *HvmTERF* genes were identified as segmental duplications. Except for *HvmTERF1* and

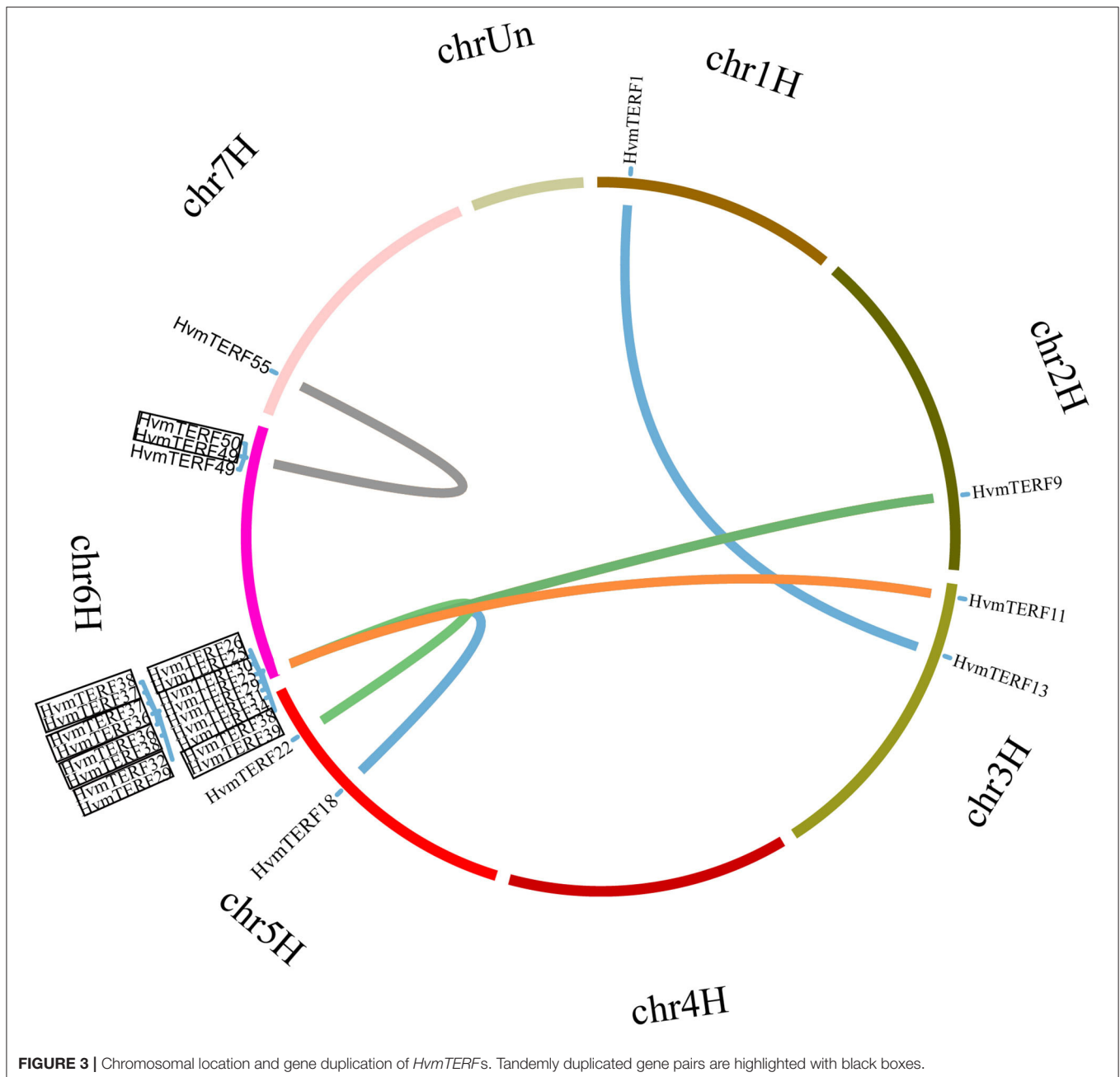
HvmTERF13, the remaining duplicated genes were clustered in subfamily IX. Remarkably, six segmentally duplicated gene pairs were associated with chromosome 6. Taken together, both tandem and segmental duplication events contributed to the expansion of *HvmTERFs*, mainly in subfamily IX.

The non-synonymous (Ka) vs. synonymous (Ks) substitution ratios for duplicated gene pairs were further calculated to estimate the evolutionary constraints acting on *HvmTERFs*. Ka/Ks >1, =1, and <1 indicate positive, neutral and purifying selection, respectively (Lynch and Conery, 2000). In this study, the Ka/Ks ratio for duplicated gene pairs ranged from 0.1516 to 0.5662, with an average value of 0.3845, suggesting that these gene pairs have experienced strong purifying selective pressure during their expansion process (Supplementary Table 7).

To further infer the evolutionary mechanisms of *HvmTERFs*, we conducted comparative ortholog analysis with five representative species, including two dicots (*A. thaliana* and *V. vinifera*) and three monocots (*B. distachyon*, *O. sativa*, and *Z. mays*) (Figure 4). The ortholog analysis resulted in 16, 18, 39, 22, and 20 gene pairs between barley and the other five species (*A. thaliana*, *V. vinifera*, *B. distachyon*, *O. sativa*, and *Z. mays*), respectively. A total of 29 *HvmTERF* genes held orthologous relationships with those in *B. distachyon*, followed by *O. sativa* (22), *Z. mays* (19), *V. vinifera* (18), and *A. thaliana* (16). Nine *HvmTERFs* (*HvmTERFs* 2, 3, 5, 8, 14, 16, 46, 58, and 59) were found to possess one-to-one relationships among the five representative species. We thus proposed that these evolutionarily conserved genes may have essential roles during plant evolution. Interestingly, 21 gene pairs composed of 12 *HvmTERFs* were only identified between barley and *B. distachyon*, *O. sativa*, and *Z. mays*, suggesting that these orthologous pairs formed after the divergence between monocotyledonous and dicotyledonous plants. The Ka/Ks ratios of the *mTERF* gene pairs were also calculated. All orthologous *mTERF* gene pairs showed Ka/Ks < 1, suggesting that these *HvmTERFs* might have been subjected to extensive purifying selective pressure (Supplementary Table 8).

Analysis of *cis*-elements and miRNA Target Sites of *HvmTERFs*

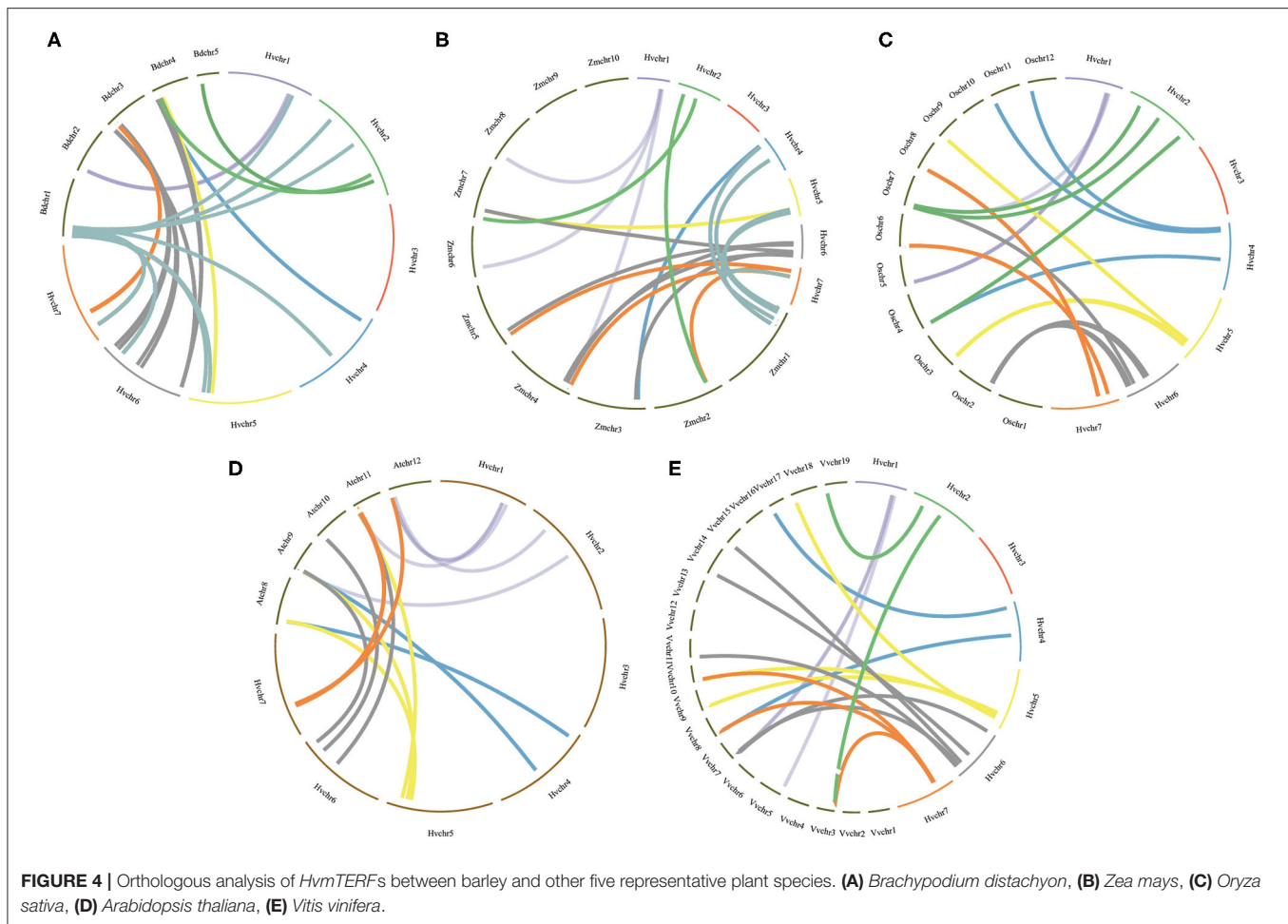
Cis-elements play vital roles in the transcriptional regulation of genes during plant growth and development as well as the plant response to biotic and abiotic stresses. By searching the PlantCARE database, a total of 40 *cis*-elements were identified and further classified into four categories. As shown in Supplementary Table 9, a total of 12 kinds of light-responsive elements were observed, accounting for the majority of the putative *cis*-elements. There were 10 hormone-responsive regulatory elements present in the *HvmTERF* promoters, such as ABRE (ABA-Responsive Elements), CGTCA-motif/TGACG-motif, TGA-element, ERE (Estrogen Response Element), and TCA-element, which were associated with ABA, methyl MeJA (Methyl Jasmonate), auxin, ethylene, and SA (Salicylic Acid), respectively. We also identified several organogenesis-related *cis*-elements, such as MSA-like (Mitosis-Specific Activator) (cell cycle regulation, 4 genes), GCN4 (General Control



Non-repressible-4) (endosperm expression, 3 genes), CAT-box (meristem expression, 32 genes), AC-I/II (xylem expression, 2 genes), and circadian (circadian control, 11 genes). Notably, five kinds of biotic and abiotic stress-related regulatory elements, including MBS (Myeloblastosis Binding Site), GC-motif, ARE (Anaerobic Response Element), LTR (Long Terminal Repeat), and Wun-motif, which responded to drought, anoxic-specific inducibility, anaerobic induction, low temperature, and wound damage, respectively, were identified in the *HvmTERF* promoter regions. Therefore, the variety and quantity of regulatory elements were present in distinct *HvmTERF* promoters, suggesting their potential functions in diverse

signal transduction pathways and various stress adaptations in barley.

To obtain preliminary insight into the miRNA-mediated posttranscriptional regulatory mechanisms, the CDSs (Coding Sequence) of *HvmTERFs* were extracted to search for miRNA target sites. The results showed that a total of 12 mTERF-miRNA pairs were identified, referring to five miRNAs targeting 10 *HvmTERFs* (**Supplementary Table 10**). Most of the miRNAs controlled the expression of *HvmTERFs* by guiding mRNA cleavage, whereas *HvmTERF34* and *HvmTERF35* were regulated by translation inhibition. *HvmTERF22* and *HvmTERF46* were targeted by miRNA6192 upstream of the



mTERF domain, whereas a total of six *HvmTERFs* were targeted by miRNA9662a-3p within the *mTERF* domain. miRNA7717b-5p and miRNA9962a-3p both targeted *HvmTERF19* through transcript cleavage. Our findings suggested that miRNA was involved in the posttranscriptional regulation of *HvmTERF*, but the actual regulatory mechanism should be validated in future molecular biology experiments.

Expression Profile Analysis of *HvmTERF* Genes

To obtain preliminary insight into tissue- and stage-specific expression profiles and elucidate the potential roles of *mTERFs* in tissue development, the transcript abundances of *HvmTERFs* in 16 different tissues or stages were obtained using Illumina RNA-seq data. As shown in **Figure 5**, *HvmTERFs* were expressed in all barley RNA-seq samples studied. The *mTERFs* were highly expressed in seedlings, leaves and developing inflorescences. *HvmTERFs* 2, 7, 8, 16, 24, 45, 46, and 58 exhibited relatively high expression levels in most of the studied tissues and stages, suggesting these genes may play an important role in barley growth and development. It is noteworthy that *HvmTERF24* ranked the most highly expressed gene with an average FPKM value of 28.58. We also

identified tissue- and stage-specific *HvmTERFs*. *HvmTERF3*, *HvmTERF5*, *HvmTERF30*, *HvmTERF35*, and *HvmTERF50* exhibited preferential expression in young inflorescences, senescing leaves, epidermis, lodicules, and senescing leaves, respectively. *HvmTERF20* was predominantly expressed in seedlings and senescing leaves, whereas *HvmTERF21* was mostly expressed in the epidermis and senescing leaves, suggesting that these genes were involved in tissue- or stage-specific development in barley. Interestingly, four *HvmTERF* genes (*HvmTERFs* 6, 26, 40, and 54) in subfamily IX exhibited almost no expression in any of the tissues and stages.

To gain comprehensive information about the functions of *HvmTERFs* in response to abiotic stresses, the expression profiles of *HvmTERFs* under cold, salt, and metal ion stresses were further investigated. The results revealed that *HvmTERFs* 15, 23, and 33 were found to be upregulated under cold treatment, whereas seven *HvmTERF* genes were downregulated (**Figure 6A**). Notably, the tissue-specific gene *HvmTERF16* (*mTERF3/SL1*, *SEEDLING LETHAL 1*) was downregulated with 2.65-fold change compared with the control. Since limited studies have been conducted on *mTERF* genes (Jiang et al., 2020), the biological function of *HvmTERFs* still need more experimental verification.

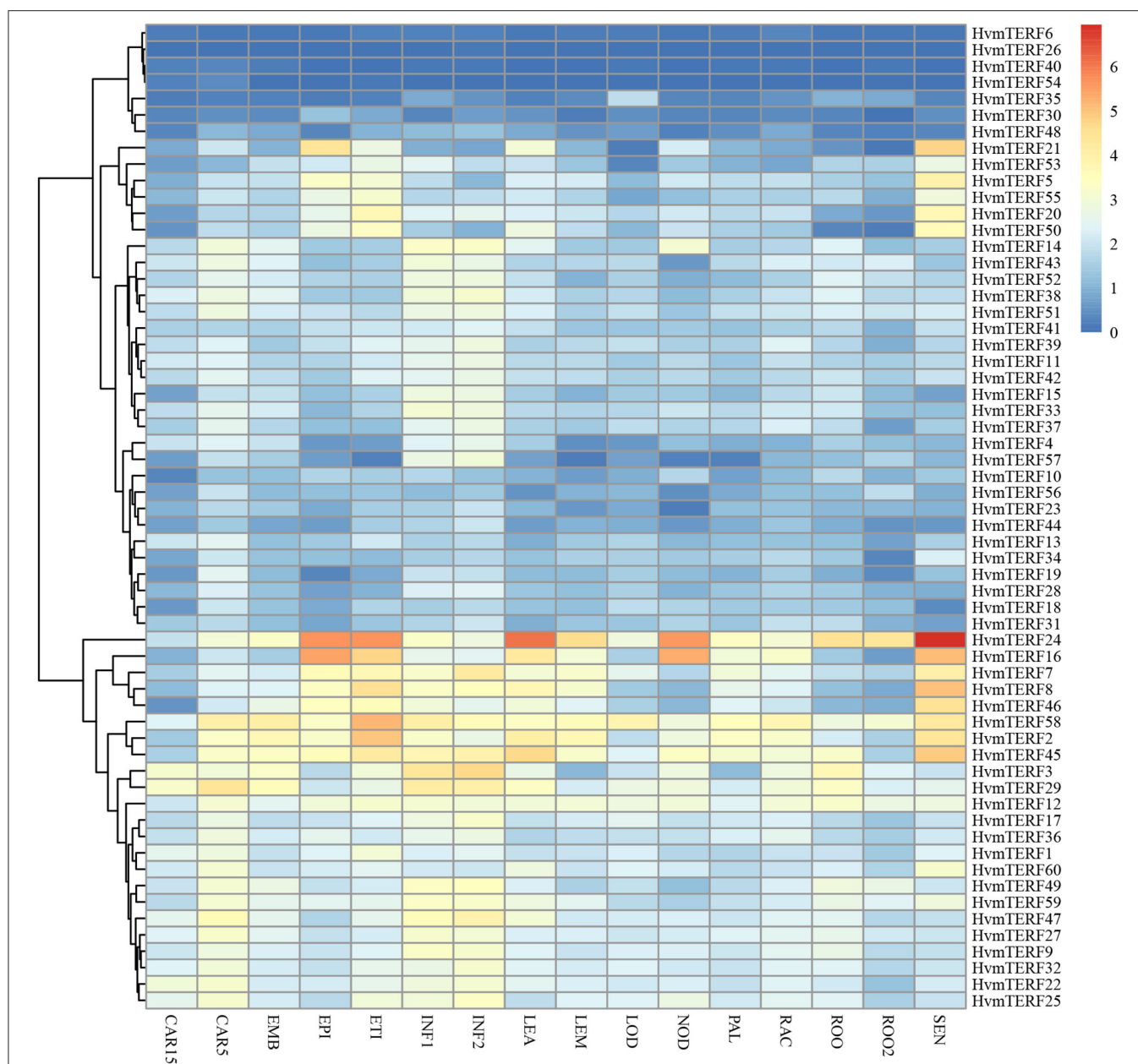


FIGURE 5 | The expression profile of *HvmTERF* genes at different tissues or stage of barley. FPKM values were normalized by $\log_2(\text{FPKM}+1)$ transform to represent color scores. CAR15/CAR5, developing grain (15/5 day after pollination); EMB, embryonic tissue (4 days); EPI, epidermal strips (4 weeks after pollination); ETI, etiolated seedling with 10 days old after planting; INF1, Young developing inflorescences with 5 mm; INF2, developing inflorescences with 1 cm; LEA, 10 cm shoots from seedlings; LEM, inflorescences, lemma (6 weeks after pollination); LOD, inflorescences, lodicule (6 weeks after pollination); NOD, developing tillers at third stem internode (6 weeks after pollination); PAL, dissected inflorescences, palea (6 weeks after pollination); RAC, inflorescences, rachis (5 weeks after pollination); ROO, roots from the seedlings with 17 and 28 days old after planting; ROO2, roots (4 weeks after pollination); SEN, senescing leaves (8 weeks after pollination).

The expression patterns of *HvmTERFs* under salt treatment were further analyzed. A total of 5, 2, and 15 *HvmTERF* genes were differentially expressed in the root meristematic zone, elongation zone, and maturation zone, respectively (**Figure 6B**). Consistent with the expression pattern under cold treatment, most (72.72%) of the differentially expressed genes, including 1 in the meristematic zone and 15 in the maturation zone,

were downregulated. Furthermore, *HvmTERF50* was upregulated in the meristematic zone and downregulated in the mature zone. *HvmTERF10* was upregulated in the elongation zone and downregulated in the mature zone. Notably, *HvmTERF21* was upregulated in both the meristematic zone (3.49-fold) and elongation zone (4.01-fold) and downregulated in the maturation zone (2.95-fold).

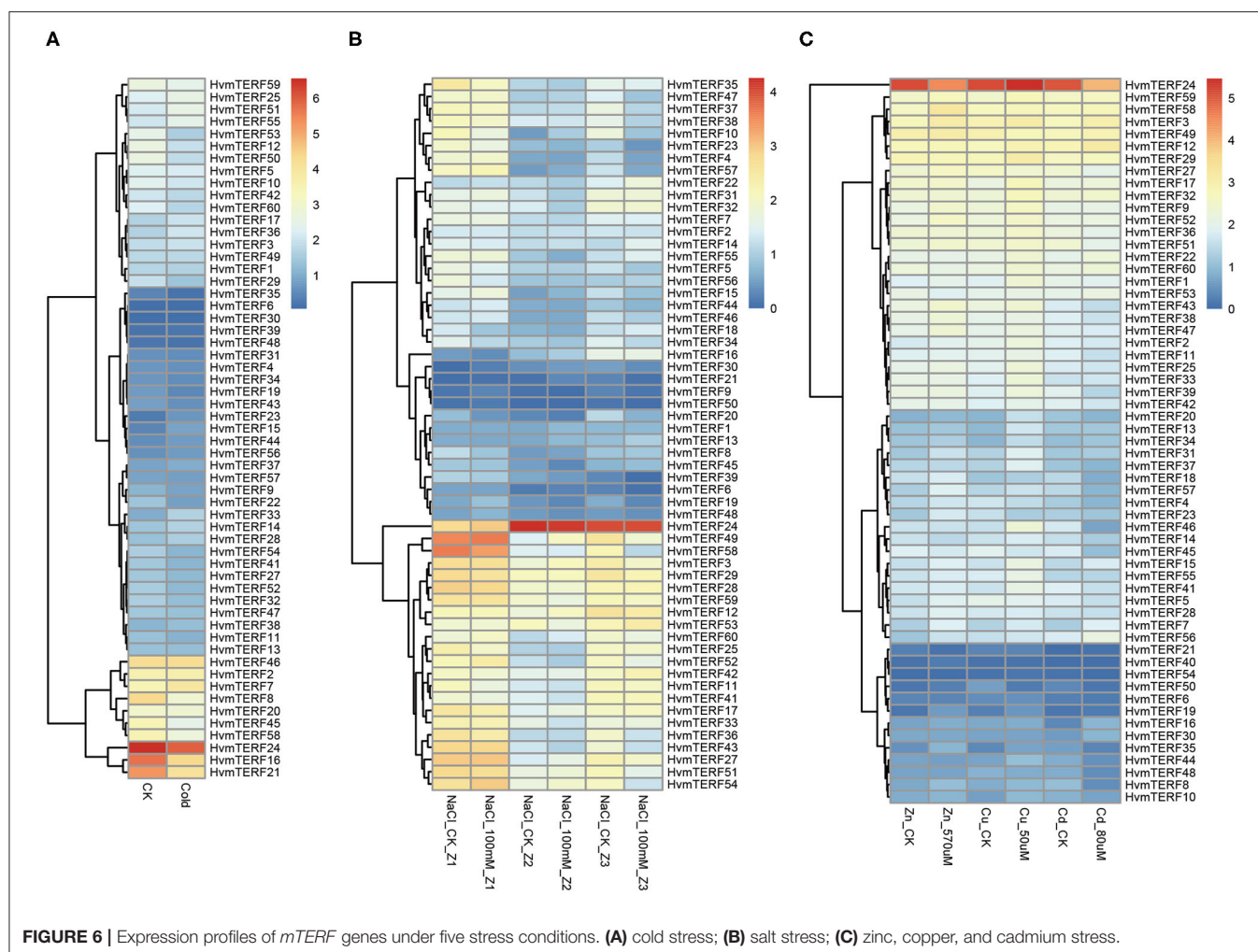


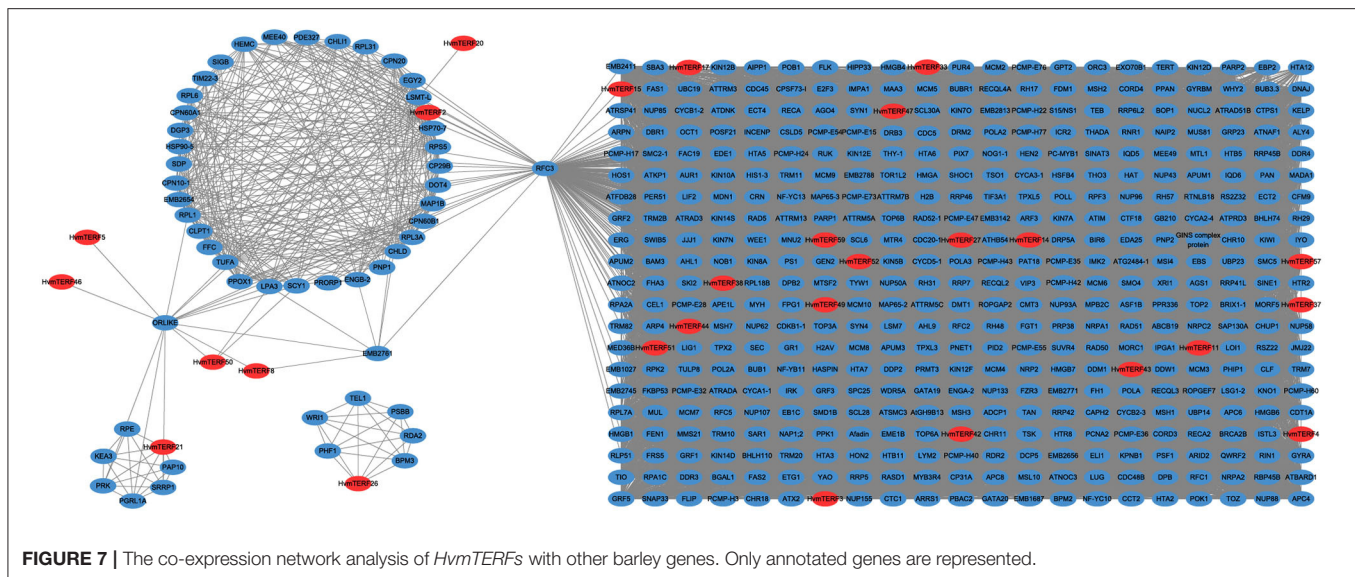
FIGURE 6 | Expression profiles of *mTERF* genes under five stress conditions. **(A)** cold stress; **(B)** salt stress; **(C)** zinc, copper, and cadmium stress.

We finally analyzed the expression profiles of *HvmTERFs* under metal ion toxicity stress. Among them, 4, 5, and 2 upregulated *HvmTERFs* were found to be zinc-, copper- and cadmium toxicity-responsive genes, and 2, 1, and 9 downregulated genes were also identified (**Figure 6C**). Remarkably, the expression levels of *HvmTERF19* under zinc and copper treatment were 4.14- and 2.45-fold higher than those of the control. *HvmTERF35* was significantly upregulated under zinc and copper treatment and downregulated under cadmium treatment. However, *HvmTERF50* was upregulated under zinc treatment and downregulated under copper and cadmium treatment. *HvmTERF21* was upregulated under cadmium treatment and downregulated under zinc treatment.

Co-expression Network Analysis Between *HvmTERFs* and Other Genes in Barley

Co-expression analysis has become an effective methodology for gene functional annotation (Wei and Chen, 2018). By using a large dataset of 148 RNA-seq samples, we attempted to construct a co-expression network of *mTERF* genes. A total of 162,373 interactions, composed of 27 *HvmTERFs* and 778

other co-expressed genes, were detected (**Figure 7**). In detail, 595 (76.48%), 260 (33.42%), 178 (22.88%), and 167 (21.47%) genes were predicted to be co-expressed with *HvmTERF57*, *HvmTERF3*, *HvmTERF15*, and *HvmTERF33*, respectively. The results suggested that these *HvmTERFs* may play central regulatory roles in the co-expression network. Interestingly, nine *HvmTERFs* were co-expressed with multiple transcription factors. For instance, 5 *B3*, 4 *GRF* (*Growth-Regulating Factor*), 3 *C3H* (*Cysteine3Histidine*), and 3 *MYB* (*Myeloblastosis*) family genes were co-expressed with 4, 4, 7, and 3 *HvmTERFs*. Transcription factors are essential regulators to repress or activate the expression of their target genes by binding to specific upstream elements (Jin et al., 2017). These results suggested that multiple transcription factors may interact with *HvmTERFs*, and further to control multitudinous growth and development processes, and response to environmental stressors in barley. Furthermore, 11 *HvmTERFs* were predicted to be co-expressed with *SPICEOSOME-ASSOCIATED PROTEIN 130* (*SAP130a*), a key gene that is required for specific spatiotemporal events during reproduction in *Arabidopsis* (Aki et al., 2011).



Gene Ontology (GO) enrichment analysis was further performed to determine the potential function of the *mTERF* co-expressed genes. The *mTERF* co-expressed genes were enriched in 155 GO terms (FDR < 0.05), including 66 biological processes, 41 cellular components, and 48 molecular functions (Supplementary Figures 4–6; Supplementary Table 11). In the molecular function category, microtubule binding (GO:0008017), nucleoside-triphosphatase activity (GO:0017111), and tubulin binding (GO:0015631) ranked as the top three enriched terms, whereas intracellular non-membrane-bound organelle (GO:0043232), non-membrane-bound organelle (GO:0043228), and nucleus (GO:0005634) were the most common terms in the cellular components category. In the biological process category, the *mTERF* co-expressed genes were implicated not only in various biological functions (e.g., GO:0006260 DNA replication metabolic process; GO:0006259 DNA metabolic process; GO:0042254 ribosome biogenesis) but also in response to diverse stresses (e.g., GO:0033554 cellular response to stress; GO:0006974 cellular response to DNA damage stimulus).

Expression of *HvmTERF* Genes in Response to Salt, Drought, Cold, and ABA Treatment via qRT-PCR

Although differentially expressed *HvmTERFs* under different stresses were obtained based on RNA-seq data, comprehensive expression patterns of *HvmTERFs* in response to various stresses and phytohormones have not been reported. To better understand the expression patterns in response to diverse stress treatments (cold, salt, heat, and ABA), 25 *HvmTERFs* were randomly selected for qRT-PCR analysis. Under salt stress, all the candidate *HvmTERFs* were downregulated after 1, 3, and 12 h of treatment. *HvmTERF23* was the most downregulated gene at the 1 h (5.93-fold), 3 h (12.89-fold), and 12 h (6.63-fold) time points (Supplementary Figure 7). Under 6 h salt treatment, nine *HvmTERFs* were found to be upregulated. Among them,

the expression level of *HvmTERF46* was dramatically increased with a 2.28-fold change value. Moreover, three upregulated *HvmTERF* genes were found at 24 h. Notably, *HvmTERF21* exhibited significantly upregulated expression levels at 6 and 24 h.

Under drought treatment, a total of 1, 1, 5, and 5 *HvmTERF* genes were upregulated at 3, 6, 12, and 24 h (Supplementary Figure 8). Among them, *HvmTERF21* showed 1.21-, 1.36-, 1.24-, and 2.58-fold changes at the 3, 6, 12, and 24 h time points, whereas *HvmTERF52* displayed 1.99- and 2.00-fold changes at the 12 and 24 h time points, respectively. The MBS *cis*-acting element involved in drought inducibility was also detected within the promoter regions of these genes (Urao et al., 1993). For example, *HvmTERFs52* possessed 2 MBS *cis*-acting elements. There were some exceptions, however, no MBS-acting element was detected in the promoter regions of *HvmTERF21*, implying this gene may have unknown elements acting in response to drought stress. Furthermore, *HvmTERFs* 2, 8, 19, 29, 43, and 49 were downregulated after drought treatment at all-time points.

Under cold treatment, we identified more upregulated *HvmTERFs* than those identified in response to salt and drought treatment, with 17, 19, 18, 17, and 20 upregulated genes after 1, 3, 6, 12, and 24 h of treatment, respectively (Supplementary Figure 9). Notably, *HvmTERFs* 9, 16, 21, 24, 25, 45, 49, and 51 were upregulated at all treated time points. The expression level of *HvmTERFs* dramatically decreased over time. The average fold change was 4.40 after 1 h of cold treatment, whereas it decreased to 1.24 after 48 h of stress, suggesting that *HvmTERFs* may mainly function in the initial response to cold injury.

The plant hormone ABA has been demonstrated to play important roles in improving the tolerance of plants to diverse stresses (Shinozaki and Yamaguchi-Shinozaki, 1997). qRT-PCR was also carried out to analyze the expression profiles of the 25 selected *HvmTERFs* after ABA treatment (Supplementary Figure 10). The heatmap revealed that *HvmTERFs* 9, 16, 21, 24, 28, and 29 exhibited upregulated expression patterns at all time points. Meanwhile, abundant

ABRE *cis*-acting elements, the major *cis*-element for ABA-responsive gene expression, were identified in the promoter regions, such as five ABRE *cis*-element for *HvmTERF28*, three for *HvmTERF16*, and three for *HvmTERF24*. The expression level of *HvmTERF46* displayed upregulated expression with a 10.16-fold change after 6 h of treatment, whereas *HvmTERF53* showed 3.89- and 3.84-fold changes after 1 and 6 h of treatment, respectively.

Nucleotide Variation, Population Structure, and Haplotype Analysis of *HvmTERF* Genes

To reveal the variation landscape of *HvmTERFs*, public resequencing data of barley were employed to detect *HvmTERF*-related SNPs. The SNP calling pipeline yielded a total of

481 *HvmTERF*-related SNPs or approximately 8.01 SNPs per gene, representing the most comprehensive variation dataset of *HvmTERFs* to date (**Supplementary Table 12**). The majority of *HvmTERF*-related SNPs (70.68%) were located in the genic region, including 159 synonymous, 133 missense, 44 intron, 3 splice region, and 1 stop-gain variant (*HvmTERF42*) (**Supplementary Table 13**). The overall transition/transversion (Ts/Tv) ratio was 2.317, with A/G (35.55%) and C/T (34.30%) ranking as the most popular allelic substitution patterns. These results indicated that there was fewer purine to purine or pyrimidine to pyrimidine mutation than there was pyrimidine to purine or purine to pyrimidine mutations.

To further investigate the genetic relationship between wild and landrace barley populations, the PCA was conducted

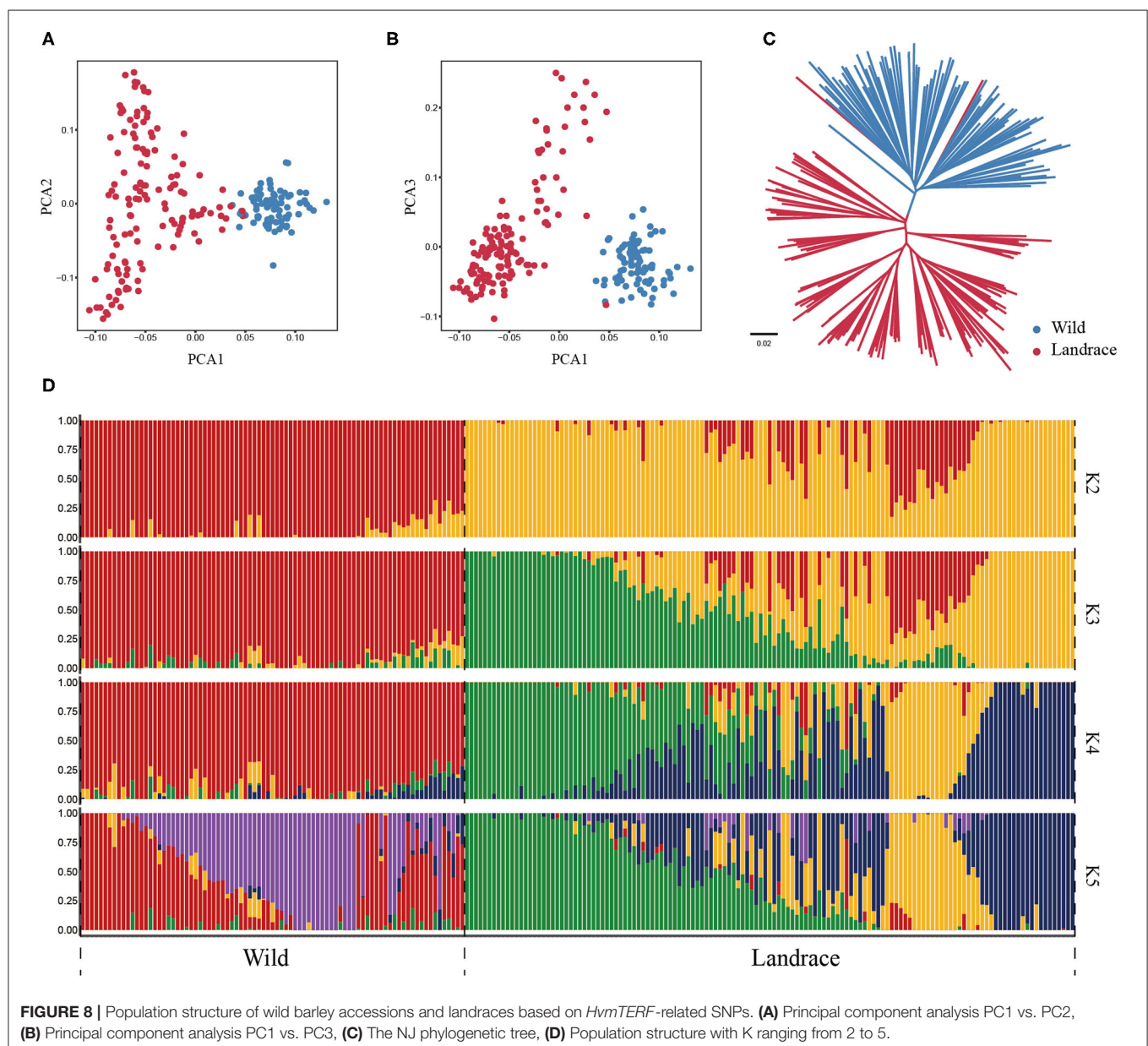


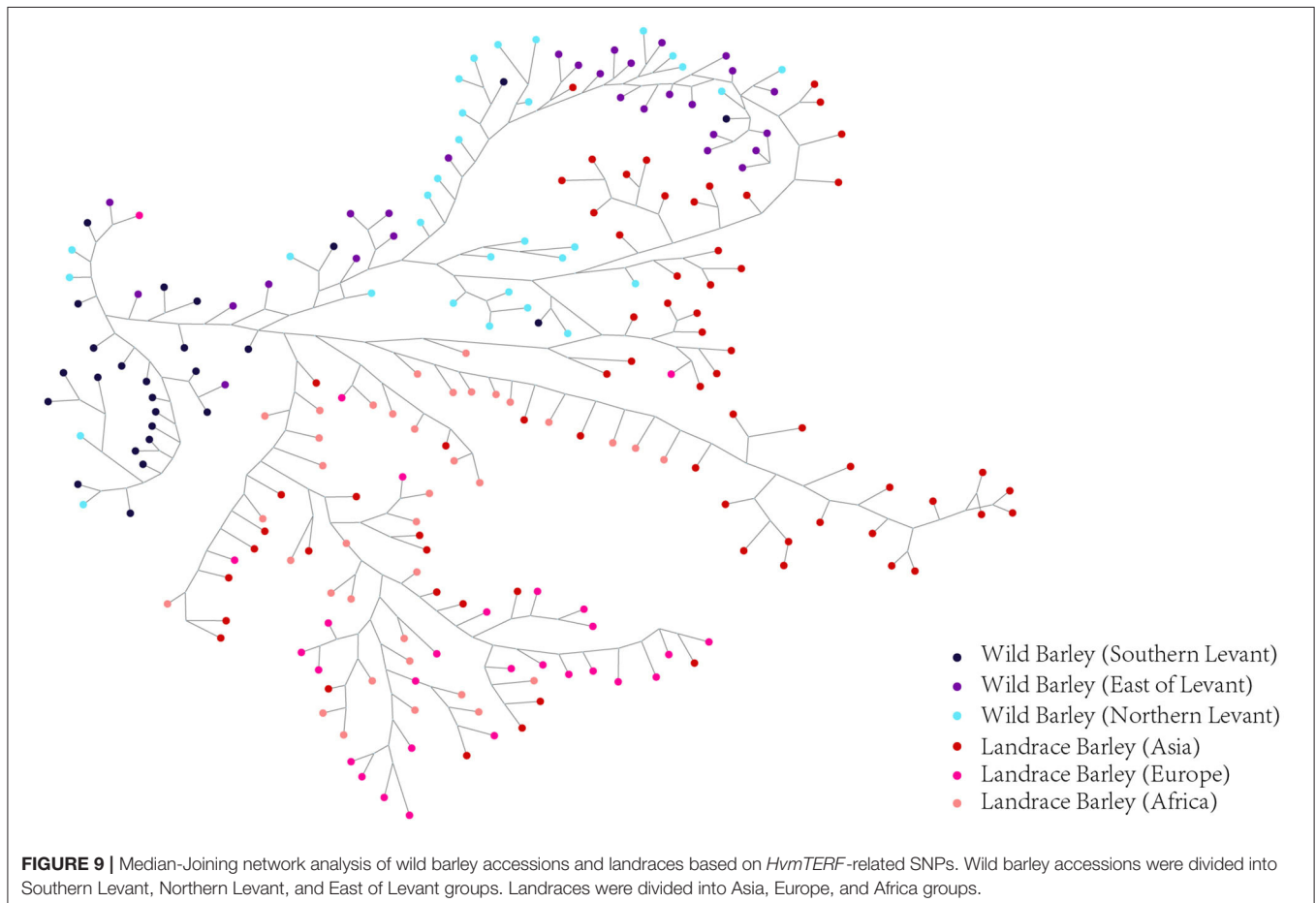
FIGURE 8 | Population structure of wild barley accessions and landraces based on *HvmTERF*-related SNPs. **(A)** Principal component analysis PC1 vs. PC2, **(B)** Principal component analysis PC1 vs. PC3, **(C)** The NJ phylogenetic tree, **(D)** Population structure with K ranging from 2 to 5.

using *HvmTERF*-related SNPs. The first eigenvector explained 23.11% of the total genetic variance and captured the biological differentiation that separated wild barley from landrace barley. The second and third eigenvectors explained 12.16 and 11.10% of the variance, respectively, and distinguished the accessions geographically (Figures 8A,B; Supplementary Table 14). The same population affinities were recovered in the phylogenetic tree with a precise accession relationship (Figure 8C). ADMIXTURE analysis also recapitulated these findings (Figure 8D). When $K = 2$, a clear separation was observed in accordance with biological differentiation between wild and landrace barley. As K increased to 5, a definite separation was presented in accordance with the geographical source. Remarkably, a certain proportion of genetic admixture between wild and landrace barley was observed, suggesting the potential domestication origin of cultivated barley and ongoing gene flow between wild and landrace barley.

Population-based nucleotide diversity (π) was estimated based on *HvmTERF*-related SNPs. The nucleotide diversity decreased only 4.5% from wild barley (0.2491) to landrace barley (0.2380), indicating that this gene family suffered a weak genetic bottleneck in the process of domestication (Supplementary Figure 11). Wright's F -statistic is an informative indicator that measures population differentiation and gene flow intensity (Wright, 1951). Populations with F_{st}

values >0.25 are considered highly differentiated (Fong et al., 2016). A relatively low F_{st} index (0.1310) was obtained between wild and landrace barley in accordance with the *HvmTERF*-related SNPs, indicating that this gene family was not subjected to strong selective pressure during barley domestication.

Haplotype dissection and comparison provide invaluable resources for understanding the evolutionary and domestication processes of important traits (Jan et al., 2019). To acquire a more precise depiction of the haplotype network, we constructed the complete haplotypes for the 220 accessions using *HvmTERF*-related SNPs. The median-joining method network generated a total of 481 *HvmTERF* haplotypes (one haplotype per accession) consisting of distinct wild and landrace groups (Figure 9). No shared haplotype between wild and landrace barley was observed in the network. The highly diverse wild accessions displayed geographical clustering patterns in terms of the Southern Levant (such as Israel and Jordan), Northern Levant (such as Syria and Turkey), and East of Levant (such as Iraq, Iran, and Central Asian counties). For landrace haplotypes, a geographical clustering pattern was obtained. However, a certain portion of accessions displayed no geographical clustering of haplotypes; for example, many shared haplotypes from Central Asia and Europe were observed in the median-joining network, suggesting a complex domestication process of *HvmTERF* in barley.



DISCUSSION

Characterization of the *mTERF* Gene Family in Barley

The *mTERF* family, firstly identified in vertebrate mitochondria, is implicated in the regulation of organellar transcription, translation, and DNA replication (Quesada, 2016; Xiong et al., 2020). With the accomplishment of various whole genome sequencing projects, the identification and characterization of *mTERF* genes have been widely conducted in diverse plant species (Tang et al., 2019). As an inbreeding, diploid and temperate cereal crop, barley is well-studied in terms of cytology, genetics and molecular biology. However, its large genome size and high transposon content have long hindered barley genome assembly projects (Schulte et al., 2009). In recent years, the first physical sequence assembly for barley and its subsequent chromosome-scale reference sequence assembly (Morex V1), as well as the improved annotated reference genome assembly (Morex V2), have formed the basis for the identification and characterization of related gene families (Jayakodi et al., 2020). In this study, we carried out a comprehensive search for putative *HvmTERFs*, and a total of 60 *HvmTERFs* were identified in the barley Morex V2 genome assembly. Compared with its previous versions, the numbers of identified *HvmTERFs* were only 40 and 51 for the barley physical sequence assembly and Morex V1 assembly, respectively (Supplementary Table 15). In addition, in the Morex V1 assembly, two *HvmTERF* members named HORVU0Hr1G013680 and HORVU0Hr1G017090 were not mapped to the reference genome. However, they were anchored to chromosome 5 (HORVU.MOREX.r2.5HG0420050, *HvmTERF*20) and chromosome 6 (HORVU.MOREX.r2.6HG0509710, *HvmTERF*49) in the Morex V2 assembly. Each of the *HvmTERFs* contains the notable conserved *mTERF* domain. No premature termination codon was found in the coding region of *HvmTERF* genes, and most of them were supported by barley ESTs, ensuring the authenticity of gene family identification. Thus, this is the first gene family identification using the Morex V2 assembly at the whole genome-wide level, and the most updated and comprehensive information on the *mTERF* gene family in barley has been obtained to date.

The improved genome assembly also provided the definite physical locations of *HvmTERFs*. The *HvmTERFs* were distributed unevenly across seven chromosomes. The maximum number of *HvmTERFs* was located on the long arm end of chromosome 6, with a total of 19 *HvmTERFs*. Most of the *HvmTERFs* were located at the distal ends of the chromosome, but they were absent from the short arm of chromosome 3 and the long arm of chromosome 5. Similar findings were reported in other barley gene families, such as the non-specific lipid transfer protein (LTP) gene family and auxin/indole-3-acetic acid (*Aux/IAA*) gene family (Zhang et al., 2019; Shi et al., 2020). A possible cause might be that the distal and proximal ends of chromosomes are more gene-rich than the middle regions of chromosomes in barley (Mayer et al., 2012). In cereals, meiotic homologous chromosome recombination is skewed toward the distal regions of the chromosomes, leading to the biased distribution of genes that are concentrated in the distal regions

(Higgins et al., 2012). The uneven distribution of recombination-rich regions ensures that there is abundant genetic diversity available to respond to various stressful conditions and dynamic environmental changes (Zhang et al., 2019).

In metazoans, the *mTERF* gene family contains four to five members (Roberti et al., 2009). In contrast, the *mTERF* gene family has expanded to approximately 30 members in land plants (Quesada, 2016). For example, there are 35 *mTERFs* in *Arabidopsis*, 33 in rice, 31 in maize, and 25 in grape. In this study, a total of 60 *HvmTERF* genes were identified in the barley reference genome, approximately two times as many as other higher plants. Given the complicated regulation of organellar genome transcription in land plants, the expansion of the *mTERF* gene family in barley could be induced by a complex mechanism. To gain insights into the evolutionary relationship within the *mTERF* gene family, we first constructed a phylogenetic tree using the *mTERF* proteins from barley, rice, and *Arabidopsis*. The *mTERF* proteins were categorized into nine subfamilies based on the classification set forth by Zhao (Zhao et al., 2014). Within the same subfamily, the gene structure and protein motif organization were highly conserved, supporting the phylogenetic analysis, and classification results. The phylogenetic tree further showed that eight subfamilies (subfamilies I–VIII) contained the *mTERF* proteins from these three species (barley, rice, and *Arabidopsis*), suggesting that these *mTERF* proteins evolved from common ancestors and then expanded independently in each species. Most of the subfamilies possessed comparable numbers of *mTERF* proteins, whereas monocot-specific subfamily IX does not contain any *mTERF* proteins from *Arabidopsis*, suggesting that subfamily IX formed after the divergence of monocots and dicots (Zhao et al., 2014). Furthermore, only 13 rice *mTERF* proteins were clustered in Group IX. In contrast, 39 *mTERF* proteins, more than half of the total *mTERFs* in barley, were assigned to this subfamily. Thus, we speculated that *HvmTERFs* within subfamily IX may have experienced noticeable expansion.

Gene duplication contributes to the expansion and evolution of gene families (Shi et al., 2020). To reveal the expansion mechanism of *HvmTERFs*, segmental and tandem duplication events were investigated. Fourteen pairs of *HvmTERFs* underwent gene duplication, including seven segmental and seven tandem duplication events. Remarkably, 17 *mTERF* genes consisting of 13 duplicated pairs contributed to the expansion of subfamily IX. Collectively, both segmental and tandem duplication contributed to the expansion of the *HvmTERF* gene family, mainly in subfamily IX, and further led to twice as many *mTERF* genes in barley as in other species. Moreover, the Ka/Ks values of all the paralogous gene pairs were <1, suggesting that they all underwent purifying selection.

HvmTERF Genes May Play Important Roles During Plant Growth, Abiotic Stress, and Phytohormone Responses

To obtain preliminary insight into the biological function of *mTERFs*, we checked the *cis*-elements in the promoter regions of *HvmTERFs*. The promoter regions contained various *cis*-elements associated with development/tissue specificity,

promoter/enhancer elements, light responses, circadian control rhythms, and external stimuli and hormone responses, suggesting that *HvmTERFs* are involved in multiple biological processes. Since *cis*-element prediction was carried out based on a bioinformatics approach, further experimental validation is also required.

In vertebrates, the biological function of *mTERFs* in the regulation of mitochondrial transcription, replication, and translation has been well-documented (Castillo et al., 2019). Although land plants possess more *mTERFs* than mammals, the functions of *mTERFs* in plants are rather limited (Kleine and Leister, 2015; Quesada, 2016). Based on their loss-of-function phenotypes, which have mainly been characterized in *Arabidopsis* and maize, *mTERFs* in land plants are required for OGE and play essential roles in plant growth and development (Ding et al., 2019). In this study, the specific spatiotemporal expression of *HvmTERFs* in different developmental stages and tissues suggested that *HvmTERFs* might potentially play a vital role in various plant growth and developmental processes. *HvmTERF2*, *HvmTERF16*, and *HvmTERF58* (orthologous to *RUGOSA2*, *SL1/mTERF3*, and *mTERF6* in *Arabidopsis*, respectively) displayed high expression levels in the studied tissues and stages. In *Arabidopsis*, these orthologous genes are essential for plant photosynthesis, mitochondrial, and chloroplastic gene expression and development, and leaf patterning and organogenesis (Quesada et al., 2011; Jiang et al., 2020). *HvmTERF24* was the most highly expressed gene in different organs. However, there is rather limited information on the functions of its orthologous gene in *Arabidopsis* (*AtmTERF12*). Recent research has only demonstrated that *AtmTERF12* is not involved in the response to salt stress (Xu et al., 2017). Several tissue- and stage-specific genes were also identified. For instance, *HvmTERF14* showed preferential expression in young inflorescences, whereas its ortholog *mTERF15* is required for the *cis*-splicing of mitochondrial *nad2* intron 3 in both *Arabidopsis* and maize and further regulates the small kernel phenotype in maize, implying that *HvmTERF14* may achieve different functions in barley compared with other species (Hsu et al., 2014; Yang et al., 2020). Homologous analysis might provide information on the role of *HvmTERFs*. However, to ascertain *HvmTERF* function still requires further detailed and extensive experimental work (Yin et al., 2021).

In contrast to animals, plants are sessile organisms that are continuously exposed to and cannot escape environmental stresses. During evolution, the expansion and diversification of gene families played important roles in plant adaptation or tolerance to environmental extremes (Xu et al., 2017). A wide range of mechanisms have evolved in plants to cope with adverse environmental stresses at the molecular level. Compared with the control, a total of three and four upregulated genes were identified under cold and salt treatment, whereas under metal poisoning stresses, a total of four zinc, five copper, and two cadmium toxicity stress-related *HvmTERFs* were identified. Since similar studies are rather scarce, further experimental validation is required. Therefore, the expression patterns of *HvmTERFs* in response to various stresses were further investigated by

qRT-PCR. Most of the qRT-PCR results were consistent with the RNA-seq results. For example, both the RNA-seq and qRT-PCR results demonstrated that *HvmTERFs* 19, 23, and 58 were downregulated in response to salt stress. By contrast, several upregulated *HvmTERFs* were also detected in response to various stresses. For instance, *HvmTERF21* was upregulated under salt and cold stress based on RNA-seq, while this gene was significantly upregulated by cold, salt, drought and ABA stress via qRT-PCR analysis. Homology analysis revealed the involvement of its orthologous gene *AtmTERF10* in salt tolerance, possibly through an ABA-mediated mechanism (Xu et al., 2017). Nonetheless, certain inconsistent expression patterns were also observed. RNA-seq data revealed that *HvmTERF7* and *HvmTERF46* were not induced by salt stress at the three root zones, while qRT-PCR analysis showed that these genes were significantly downregulated at 1, 3, 12, and 24 h and significantly upregulated at 6 h under salt treatment. Previous studies reported that *mTERF9* (orthologous to *HvmTERF7*) and *MAD1/mTERF5* (orthologous to *HvmTERF46*) contributed to salt tolerance in *Arabidopsis* (Robles et al., 2012, 2015; Núñez-Delegido et al., 2020). The inconsistent results between RNA-seq and qRT-PCR may be due to several putative reasons. First, the different barley varieties were used in the two experiments. The barley cultivar Clipper was used in RNA-seq, whereas the cultivar Morex was the experimental materials in qRT-PCR. Second, the plant materials were not exactly the same. The materials for qRT-PCR were roots, whereas the materials for RNA-seq were both roots and leaves. Third, the expression level of qRT-PCR was calculated based on the $2^{-\Delta\Delta CT}$ method, and the expression level of RNA-seq was estimated by FPKM. These two different calculation methods could not ensure that all the results are consistent. In brief, these results provide candidates for further functional investigation of *mTERF* genes in barley as well as in other cereal crops.

In addition, there was no correlation between expression patterns and phylogenetic relationships. The fate of *HvmTERF* genes from the same gene family could be described as neofunctionalization during expansion. For example, in subfamily IV, *HvmTERFs* 2, 8, and 46 showed relatively high expression in various tissues and developmental stages, whereas *HvmTERF20* exhibited preferential expression in developing inflorescences and senescing leaves, and *HvmTERF51* was not expressed in most of the developmental stages and tissues. Highly diverse expression patterns were also observed in subfamily IX, the most expanded subfamily. In addition, a divergent expression profile was investigated even for duplicated gene pairs. The duplicated genes *HvmTERF9* and 34 showed divergent spatiotemporal expression patterns. *HvmTERF34* was induced by copper, whereas its paralog, *HvmTERF9*, was significantly upregulated in the root meristematic zone under salt treatment. Similar patterns were also observed in other phylogenetic groups. These results suggested that close phylogenetic relationships are not essential for similar expression profiles, which was consistent with previous reports on other barley gene families (Li et al., 2019).

Nucleotide Diversity Analysis Indicated That *HvmTERF* Genes Experienced a Weak Bottleneck During Barley Domestication

Domestication is a plant-animal coevolution process that occurs when wild species are exposed to new selective environments associated with human cultivation and use, leading to morphological and physical changes that distinguish domesticated taxa from their wild ancestors (Purugganan and Fuller, 2009; Purugganan, 2019). Cultivated barley, domesticated from its progenitor wild barley (*Hordeum vulgare* ssp. *spontaneum*), has experienced a series of genetic changes that have caused differences in plant architecture and growth habits, collectively called the domestication syndrome (Hammer, 1984; Doebley et al., 2006). Domestication of barley resulted in a concomitant bottleneck that reduced nucleotide diversity in alleles (Haas et al., 2020). However, little is known about the changes in *mTERFs* resulting from domestication in barley. In this study, 481 SNPs were identified from 60 *HvmTERF* genes across 220 wild and landrace barley accessions. The SNPs were distributed unevenly along the genomic sequence, including a total of 292 exon and 44 intron variations, which was consistent with a previous study showing higher polymorphism of SNPs in exon regions than in intron regions (Lu et al., 2019) but opposite to observations in other studies (Uçarli et al., 2016; Xia et al., 2017). The PCA, admixture, and phylogenetic analyses clearly divided all the accessions into landraces and wild barley accessions and further distinguished them geographically. We further examined the geographical distribution of these haplotypes and found that the wild barley populations from the Northern Levant and East of Levant regions appeared to contribute most directly to the genetic composition of Asian landraces, while Southern Levant barley populations made a great contribution to African and European landraces. The genetic constitution of barley landraces indicated multiple origins from wild progenitor populations that resulted in the initial domestication and subsequent migration of early agriculturalists along the axes of the Afro-Eurasian world (Poets et al., 2015). Although multiple wild populations provided the basis for the genetic constitution of landraces, the broad geographic range of landraces also showed various regional correlations.

Domestication also resulted in a concomitant bottleneck that reduced sequence diversity across many genes (Haas et al., 2020). The nucleotide diversity of wild accessions was relatively higher than that of landrace accessions, with a total decrease of ~4.5%. Compared with the previous study, the average reduction in nucleotide diversity was 27% from wild barley accessions to landraces across the genome (Russell et al., 2016). The significantly lower nucleotide diversity loss passing from wild barley accessions to landraces in this study indicated that the *HvmTERF* gene family might have suffered simple bottleneck effects, rather than selection, in the process of barley domestication. This result was also verified by the relatively low *Fst* index. No significant divergence occurred between wild barley accessions and landraces regarding *HvmTERFs*. One plausible reason is that the hitchhiking effect reduced the nucleotide

diversity of the linked loci associated with domestication (Kilian et al., 2006).

As in other plants, *mTERFs* are characterized as evolutionarily conserved and fundamentally universal signaling pathways. However, the comprehensive characterization of barley *mTERF* gene family remains to be elucidated in detail. Our data on the physicochemical properties, phylogenetic relationships, gene structures, conserved motifs, expansion patterns, expression profiles, and genetic variations will provide essential clues for investigating the biological functions and evolutionary history of *mTERF* gene family in barley.

CONCLUSION

In this study, a total of 60 *mTERF* genes were identified in barley, about two times as many as those in *Arabidopsis* and rice. Phylogenetic analysis categorized these genes into nine subfamilies, with approximately half of them assigned to subfamily IX, which was supported by exon-intron structure and conserved motif analyses. Both segmental and tandem duplications contributed significantly to the expansion of *HvmTERFs*, mainly in subfamily IX. *Cis*-acting regulatory element, expression profile and co-expression network analyses suggested that *HvmTERFs* might be involved in the development process, tolerance to diverse stresses and response to plant hormones. qRT-PCR analysis also revealed that *HvmTERF21* and *HvmTERF23* were significant induced by several abiotic stresses and/or phytohormone treatment, and these genes could be considered candidates for further functional studies. Finally, genetic variation analysis demonstrated that *HvmTERFs* may have experienced a weak genetic bottleneck, rather than selection, during the domestication process from wild to landrace barley. Taken together, our findings will not only provide a solid foundation for further evolutionary analysis but also facilitate the functional study of *HvmTERFs* and the molecular breeding of barley.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

XW, FW, and LC designed and supervised the project. TL, WP, YY, and YLiu performed the data analysis. WP and YLiu performed the experiments. TL, YLi, and LC drafted the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Natural Science Foundation of China (Grant No. 32060458), Jiangxi Natural Science Foundation (Grant No. 20202BAB215002), and the

Science and Technology Research Project of Jiangxi Provincial Department of Education (Grant No. GJJ180241). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.684619/full#supplementary-material>

Supplementary Figure 1 | Schematic geographic distribution of 220 wild barley and landraces accessions from different regions.

Supplementary Figure 2 | Sequence characteristics and conserved amino acids ratio of mTERF proteins in four species. (A) Barley, (B) Rice, (C) *Arabidopsis*, (D) Maize.

Supplementary Figure 3 | Distribution of *HvmTERF* genes on barley chromosomes. Chromosome numbers are shown at the top of each chromosome. The scale (Mb) is indicated on the left.

Supplementary Figure 4 | GO enrichment (biological process) of *HvmTERF* co-expressed genes in the WGCNA network.

Supplementary Figure 5 | GO enrichment (cellular component) of *HvmTERF* co-expressed genes in the WGCNA network.

Supplementary Figure 6 | GO enrichment (molecular function) of *HvmTERF* co-expressed genes in the WGCNA network.

Supplementary Figure 7 | The expression analysis of *HvmTERF* genes in response to salt. Error bar represent the standard error of the mean. One asterisk (*) indicates 0.05 significance level. Double asterisk (**) indicates 0.01 significance level.

Supplementary Figure 8 | The expression analysis of *HvmTERF* genes in response to drought. Error bar represent the standard error of the mean. One asterisk (*) indicates 0.05 significance level. Double asterisk (**) indicates 0.01 significance level.

Supplementary Figure 9 | The expression analysis of *HvmTERF* genes in response to cold. Error bar represent the standard error of the mean. One asterisk (*) indicates 0.05 significance level. Double asterisk (**) indicates 0.01 significance level.

Supplementary Figure 10 | The expression analysis of *HvmTERF* genes in response to ABA. Error bar represent the standard error of the mean. One asterisk (*) indicates 0.05 significance level. Double asterisk (**) indicates 0.01 significance level.

Supplementary Figure 11 | Nucleotide diversity of wild barley and landraces accessions. The horizontal line inside the box corresponds to the median of this distribution, the bottom and top of the box are the first and third quartiles and data points outside the whiskers can be considered as outliers. The nucleotide diversity was calculated based on *HvmTERF*-related SNPs.

Supplementary Table 1 | Accession number and samples information of RNA-seq data used in this study.

Supplementary Table 2 | The primers used in the qRT-PCR.

Supplementary Table 3 | Summary information of the 220 wild barley accessions and landraces.

Supplementary Table 4 | Characteristic features of *mTERF* gene family in barley.

Supplementary Table 5 | Verification of the mTERF domains using the NCBI-CDD, SMART, HMMER, and InterPro online tools.

Supplementary Table 6 | Consensus motifs analysis of mTERF domains in different species.

Supplementary Table 7 | The segmental and tandem duplication events of *HvmTERFs*.

Supplementary Table 8 | The Ka/Ks ratios and estimated divergence time between barley and other five species (*A. thaliana*, *B. distachyon*, *O. sativa*, *Vitis vinifera*, *Zea mays*).

Supplementary Table 9 | Characteristics of *cis*-acting regulatory elements in the promoter regions of the *HvmTERF* genes.

Supplementary Table 10 | List of putative miRNAs targeted to *HvmTERF* genes identified by psRNATarget online tool.

Supplementary Table 11 | GO enrichment of *HvmTERF* co-expressed genes.

Supplementary Table 12 | The genetic variation information (VCF format) of *HvmTERF*-related SNPs across the 220 wild barley accessions and landraces.

Supplementary Table 13 | Distribution of SNPs within various genomic regions.

Supplementary Table 14 | Tracy-Widom (TW) statistics and *P*-values for the first 10 eigenvalues in the PCA of *HvmTERF*-related SNPs.

Supplementary Table 15 | The number of *HvmTERFs* from different barley assemblies.

REFERENCES

- Aki, S., Nakai, H., Aoyama, T., Oka, A., and Tsuge, T. (2011). AtSAP130/AtSF3b-3 function is required for reproduction in *Arabidopsis thaliana*. *Plant Cell Physiol.* 52, 1330–1339. doi: 10.1093/pcp/pcr077
- Andersson, D. C., Fauconnier, J., Park, C. B., Zhang, S. J., Thireau, J., Ivarsson, N., et al. (2011). Enhanced cardiomyocyte Ca(2+) cycling precedes terminal AV-block in mitochondrial cardiomyopathy Mterf3 KO mice. *Antioxid. Redox Signal.* 15, 2455–2464. doi: 10.1089/ars.2011.3915
- Binder, S., and Brennicke, A. (2003). Gene expression in plant mitochondria: transcriptional and post-transcriptional control. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 181–188; discussion 188–189. doi: 10.1098/rstb.2002.1179
- Cámara, Y., Asin-Cayuela, J., Park, C. B., Metodiev, M. D., Shi, Y., Ruzzenente, B., et al. (2011). MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome. *Cell Metab.* 13, 527–539. doi: 10.1016/j.cmet.2011.04.002
- Castillo, A., Vilà, M., Pedriza, I., Pardo, R., Cámara, Y., Martín, E., et al. (2019). Adipocyte MTERF4 regulates non-shivering adaptive thermogenesis and sympathetic-dependent glucose homeostasis. *Biochim. Biophys. Acta Mol. Basis Dis.* 1865, 1298–1312. doi: 10.1016/j.bbdis.2019.01.025
- Chen, L., Hu, W., Tan, S., Wang, M., Ma, Z., Zhou, S., et al. (2012). Genome-wide identification and analysis of MAPK and MAPKK gene families in *Brachypodium distachyon*. *PLoS ONE* 7:e46744. doi: 10.1371/journal.pone.0046744
- Ding, S., Zhang, Y., Hu, Z., Huang, X., Zhang, B., Lu, Q., et al. (2019). mTERF5 acts as a transcriptional pausing factor to positively regulate transcription of chloroplast psbEFLJ. *Mol. Plant* 12, 1259–1277. doi: 10.1016/j.molp.2019.05.007
- Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321. doi: 10.1016/j.cell.2006.12.006
- Fong, M. Y., Rashdi, S. A., Yusof, R., and Lau, Y. L. (2016). Genetic diversity, natural selection and haplotype grouping of *Plasmodium knowlesi* gamma protein region II (PkyRII): comparison with the duffy binding protein (PkDBPαRII). *PLoS ONE* 11:e0155627. doi: 10.1371/journal.pone.0155627
- Gray, M. W. (2012). Mitochondrial evolution. *Cold Spring Harb. Perspect. Biol.* 4:a011403. doi: 10.1101/cshperspect.a011403
- Gu, Z., Cavalcanti, A., Chen, F. C., Bouman, P., and Li, W. H. (2002). Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol. Biol. Evol.* 19, 256–262. doi: 10.1093/oxfordjournals.molbev.a004079
- Haas, M., Himmelbach, A., and Mascher, M. (2020). The contribution of cis- and trans-acting variants to gene regulation in wild and domesticated barley under cold stress and control conditions. *J. Exp. Bot.* 71, 2573–2584. doi: 10.1093/jxb/eraa036

- Hammani, K., and Barkan, A. (2014). An mTERF domain protein functions in group II intron splicing in maize chloroplasts. *Nucleic Acids Res.* 42, 5033–5042. doi: 10.1093/nar/gku112
- Hammer, K. (1984). Das domestikationssyndrom. *Die Kulturpf.* 32, 11–34. doi: 10.1007/BF02098682
- Higgins, J. D., Perry, R. M., Barakate, A., Ramsay, L., Waugh, R., Halpin, C., et al. (2012). Spatiotemporal asymmetry of the meiotic program underlies the predominantly distal distribution of meiotic crossovers in barley. *Plant Cell* 24, 4096–4109. doi: 10.1105/tpc.112.102483
- Hsu, Y. W., Wang, H. J., Hsieh, M. H., Hsieh, H. L., and Jauh, G. Y. (2014). Arabidopsis mTERF15 is required for mitochondrial nad2 intron 3 splicing and functional complex I activity. *PLoS ONE* 9:e112360. doi: 10.1371/journal.pone.0112360
- Huang, M., Friso, G., Nishimura, K., Qu, X., Olinares, P. D., Majeran, W., et al. (2013). Construction of plastid reference proteomes for maize and Arabidopsis and evaluation of their orthologous relationships; the concept of orthoproteomics. *J. Proteome Res.* 12, 491–504. doi: 10.1021/pr300952g
- Huang, W., Yu, M., Jiao, Y., Ma, J., Ma, M., Wang, Z., et al. (2011). Mitochondrial transcription termination factor 2 binds to entire mitochondrial DNA and negatively regulates mitochondrial gene expression. *Acta Biochim. Biophys. Sin.* 43, 472–479. doi: 10.1093/abbs/gmr035
- Jan, H. U., Guan, M., Yao, M., Liu, W., Wei, D., Abbadi, A., et al. (2019). Genome-wide haplotype analysis improves trait predictions in *Brassica napus* hybrids. *Plant Sci.* 283, 157–164. doi: 10.1016/j.plantsci.2019.02.007
- Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V. S., Gundlach, H., Monat, C., et al. (2020). The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* 588, 284–289. doi: 10.1038/s41586-020-2947-8
- Jiang, D., Chen, J., Zhang, Z., and Hou, X. (2021). mitochondrial transcription termination factor 27 is required for salt tolerance in *Arabidopsis thaliana*. *Int. J. Mol. Sci.* 22. doi: 10.3390/ijms22031466
- Jiang, D., Tang, R., Shi, Y., Ke, X., Wang, Y., Che, Y., et al. (2020). Arabidopsis seedling lethal 1 interacting with plastid-encoded RNA polymerase complex proteins is essential for chloroplast development. *Front. Plant Sci.* 11:602782. doi: 10.3389/fpls.2020.602782
- Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. doi: 10.1093/nar/gkw982
- Kilian, B., Ozkan, H., Kohl, J., von Haeseler, A., Barale, F., Deus, O., et al. (2006). Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. *Mol. Genet. Genom.* 276, 230–241. doi: 10.1007/s00438-006-0136-6
- Kim, M., Lee, U., Small, I., des Franches-Small, C. C., and Vierling, E. (2012). Mutations in an Arabidopsis mitochondrial transcription termination factor-related protein enhance thermotolerance in the absence of the major molecular chaperone HSP101. *Plant Cell* 24, 3349–3365. doi: 10.1105/tpc.112.101006
- Kleine, T. (2012). Arabidopsis thaliana mTERF proteins: evolution and functional classification. *Front. Plant Sci.* 3:233. doi: 10.3389/fpls.2012.00233
- Kleine, T., and Leister, D. (2015). Emerging functions of mammalian and plant mTERFs. *Biochim. Biophys. Acta* 1847, 786–797. doi: 10.1016/j.bbabi.2014.12.009
- Kruse, B., Narasimhan, N., and Attardi, G. (1989). Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination. *Cell* 58, 391–397. doi: 10.1016/0092-8674(89)90853-2
- Kühn, K., Böhne, A. V., Liere, K., Weihe, A., and Börner, T. (2007). Arabidopsis phage-type RNA polymerases: accurate *in vitro* transcription of organellar genes. *Plant Cell* 19, 959–971. doi: 10.1105/tpc.106.046839
- Lang, B. F., Gray, M. W., and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* 33, 351–397. doi: 10.1146/annurev.genet.33.1.351
- Lee, C. P., Taylor, N. L., and Millar, A. H. (2013). Recent advances in the composition and heterogeneity of the Arabidopsis mitochondrial proteome. *Front. Plant Sci.* 4:4. doi: 10.3389/fpls.2013.00004
- Lee, K., Leister, D., and Kleine, T. (2021). Arabidopsis mitochondrial transcription termination factor mTERF2 promotes splicing of group IIB introns. *Cells* 10:315. doi: 10.3390/cells10020315
- Leister, D., and Kleine, T. (2020). Extending the repertoire of mTERF proteins with functions in organellar gene expression. *Mol. Plant* 13, 817–819. doi: 10.1016/j.molp.2020.04.003
- Li, Y., Xiong, H., Cuo, D., Wu, X., and Duan, R. (2019). Genome-wide characterization and expression profiling of the relation of the HD-Zip gene family to abiotic stress in barley (*Hordeum vulgare* L.). *Plant Physiol. Biochem.* 141, 250–258. doi: 10.1016/j.plaphy.2019.05.026
- Liere, K., Weihe, A., and Börner, T. (2011). The transcription machineries of plant mitochondria and chloroplasts: composition, function, and regulation. *J. Plant Physiol.* 168, 1345–1360. doi: 10.1016/j.jplph.2011.01.005
- Linder, T., Park, C. B., Asin-Cayuela, J., Pellegrini, M., Larsson, N. G., Falkenberg, M., et al. (2005). A family of putative transcription termination factors shared amongst metazoans and plants. *Curr. Genet.* 48, 265–269. doi: 10.1007/s00294-005-0022-5
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Lu, X., Fang, Y., Tian, B., Tong, T., Wang, J., Wang, H., et al. (2019). Genetic variation of HvXYN1 associated with endoxylanase activity and TAX content in barley (*Hordeum vulgare* L.). *BMC Plant Biol.* 19:170. doi: 10.1186/s12870-019-1747-5
- Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151
- Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544, 427–433. doi: 10.1038/nature22043
- Mayer, K. F., Waugh, R., Brown, J. W., Schulman, A., Langridge, P., Platzer, M., et al. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491, 711–716. doi: 10.1038/nature11543
- Meskauskiene, R., Würsch, M., Laloi, C., Vidi, P. A., Coll, N. S., Kessler, F., et al. (2009). A mutation in the Arabidopsis mTERF-related plastid protein SOLDAT10 activates retrograde signaling and suppresses (1)O(2)-induced cell death. *Plant J.* 60, 399–410. doi: 10.1111/j.1365-313X.2009.03965.x
- Méteignier, L. V., Ghandour, R., Meierhoff, K., Zimmerman, A., Chicher, J., Baumberger, N., et al. (2020). The Arabidopsis mTERF-repeat MDA1 protein plays a dual function in transcription and stabilization of specific chloroplast transcripts within the psbE and ndhH operons. *New Phytol.* 227, 1376–1391. doi: 10.1111/nph.16625
- Monat, C., Padmarasu, S., Lux, T., Wicker, T., Gundlach, H., Himmelbach, A., et al. (2019). TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools. *Genome Biol.* 20:284. doi: 10.1186/s13059-019-1899-5
- Núñez-Delegido, E., Robles, P., Ferrández-Ayala, A., and Quesada, V. (2020). Functional analysis of mTERF5 and mTERF9 contribution to salt tolerance, plastid gene expression and retrograde signalling in *Arabidopsis thaliana*. *Plant Biol.* 22, 459–471. doi: 10.1111/plb.13084
- Pan, Z., Ren, X., Zhao, H., Liu, L., Tan, Z., and Qiu, F. (2019). A mitochondrial transcription termination factor, ZmSmk3, is required for nad1 Intron4 and nad4 intron1 splicing and kernel development in maize. *G3* 9, 2677–2686. doi: 10.1534/g3.119.400265
- Park, C. B., Asin-Cayuela, J., Cámara, Y., Shi, Y., Pellegrini, M., Gaspari, M., et al. (2007). MTERF3 is a negative regulator of mammalian mtDNA transcription. *Cell* 130, 273–285. doi: 10.1016/j.cell.2007.05.046
- Pellegrini, M., Asin-Cayuela, J., Erdjument-Bromage, H., Tempst, P., Larsson, N. G., and Gustafsson, C. M. (2009). MTERF2 is a nucleoid component in mammalian mitochondria. *Biochim. Biophys. Acta* 1787, 296–302. doi: 10.1016/j.bbabi.2009.01.018
- Pfannschmidt, T., Blavillain, R., Merendino, L., Courtois, F., Chevalier, F., Liebers, M., et al. (2015). Plastid RNA polymerases: orchestration of enzymes with different evolutionary origins controls chloroplast biogenesis during the plant life cycle. *J. Exp. Bot.* 66, 6957–6973. doi: 10.1093/jxb/erv415
- Poets, A. M., Fang, Z., Clegg, M. T., and Morrell, P. L. (2015). Barley landraces are characterized by geographically heterogeneous genomic origins. *Genome Biol.* 16:173. doi: 10.1186/s13059-015-0712-3
- Purugganan, M. D. (2019). Evolutionary insights into the nature of plant domestication. *Curr. Biol.* 29, R705–R714. doi: 10.1016/j.cub.2019.05.053
- Purugganan, M. D., and Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature* 457, 843–848. doi: 10.1038/nature07895

- Quesada, V. (2016). The roles of mitochondrial transcription termination factors (MTERFs) in plants. *Physiol. Plant.* 157, 389–399. doi: 10.1111/ppl.12416
- Quesada, V., Sarmiento-Mañú, R., González-Bayón, R., Hricová, A., Pérez-Marcos, R., Graciá-Martínez, E., et al. (2011). Arabidopsis RUGOSA2 encodes an mTERF family member required for mitochondrion, chloroplast and leaf development. *Plant J.* 68, 738–753. doi: 10.1111/j.1365-313X.2011.04726.x
- Roberti, M., Bruni, F., Polosa, P. L., Gadaleta, M. N., and Cantatore, P. (2006). The Drosophila termination factor DmTTF regulates *in vivo* mitochondrial transcription. *Nucleic Acids Res.* 34, 2109–2116. doi: 10.1093/nar/gkl181
- Roberti, M., Polosa, P. L., Bruni, F., Manzari, C., Deceglie, S., Gadaleta, M. N., et al. (2009). The MTERF family proteins: mitochondrial transcription regulators and beyond. *Biochim. Biophys. Acta* 1787, 303–311. doi: 10.1016/j.bbabi.2009.01.013
- Robles, P., Micol, J. L., and Quesada, V. (2012). Arabidopsis MDA1, a nuclear-encoded protein, functions in chloroplast development and abiotic stress responses. *PLoS ONE* 7:e42924. doi: 10.1371/journal.pone.0042924
- Robles, P., Micol, J. L., and Quesada, V. (2015). Mutations in the plant-conserved MTERF9 alter chloroplast gene expression, development and tolerance to abiotic stress in *Arabidopsis thaliana*. *Physiol. Plant.* 154, 297–313. doi: 10.1111/ppl.12307
- Robles, P., Navarro-Cartagena, S., Ferrández-Ayela, A., Núñez-Delegido, E., and Quesada, V. (2018a). The characterization of Arabidopsis mterf6 mutants reveals a new role for mTERF6 in tolerance to abiotic stress. *Int. J. Mol. Sci.* 19:2388. doi: 10.3390/ijms19082388
- Robles, P., Núñez-Delegido, E., Ferrández-Ayela, A., Sarmiento-Mañú, R., Micol, J. L., and Quesada, V. (2018b). Arabidopsis mTERF6 is required for leaf patterning. *Plant Sci.* 266, 117–129. doi: 10.1016/j.plantsci.2017.11.003
- Robles, P., and Quesada, V. (2021). Research progress in the molecular functions of plant mTERF proteins. *Cells* 10:205. doi: 10.3390/cells10020205
- Romani, I., Manavski, N., Morosetti, A., Tadini, L., Maier, S., Kühn, K., et al. (2015). A member of the arabidopsis mitochondrial transcription termination factor family is required for maturation of chloroplast transfer RNA^{Ile}(GAU). *Plant Physiol.* 169, 627–646. doi: 10.1104/pp.15.00964
- Russell, J., Mascher, M., Dawson, I. K., Kyriakidis, S., Calixto, C., Freund, F., et al. (2016). Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. *Nat. Genet.* 48, 1024–1030. doi: 10.1038/ng.3612
- Schulte, D., Close, T. J., Graner, A., Langridge, P., Matsumoto, T., Muehlbauer, G., et al. (2009). The international barley sequencing consortium—at the threshold of efficient access to the barley genome. *Plant Physiol.* 149, 142–147. doi: 10.1104/pp.108.128967
- Shi, Q., Zhang, Y., To, V. T., Shi, J., Zhang, D., and Cai, W. (2020). Genome-wide characterization and expression analyses of the auxin/indole-3-acetic acid (Aux/IAA) gene family in barley (*Hordeum vulgare* L.). *Sci. Rep.* 10:10242. doi: 10.1038/s41598-020-66860-7
- Shinozaki, K., and Yamaguchi-Shinozaki, K. (1997). Gene expression and signal transduction in water-stress response. *Plant Physiol.* 115, 327–334. doi: 10.1104/pp.115.2.327
- Spähr, H., Habermann, B., Gustafsson, C. M., Larsson, N. G., and Hallberg, B. M. (2012). Structure of the human MTERF4-NSUN4 protein complex that regulates mitochondrial ribosome biogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15253–15258. doi: 10.1073/pnas.1210688109
- Tang, B., Xie, L., Yi, T., Lv, J., Yang, H., Cheng, X., et al. (2019). Genome-wide identification and characterization of the mitochondrial transcription termination factors (mTERFs) in *Capsicum annuum* L. *Int. J. Mol. Sci.* 21:269. doi: 10.3390/ijms21010269
- Terzioglu, M., Ruzzenente, B., Harmel, J., Mourier, A., Jemt, E., López, M. D., et al. (2013). MTERF1 binds mtDNA to prevent transcriptional interference at the light-strand promoter but is dispensable for rRNA gene transcription regulation. *Cell Metab.* 17, 618–626. doi: 10.1016/j.cmet.2013.03.006
- Uçarlı, C., McGuffin, L. J., Çaputlu, S., Aravena, A., and Gürel, F. (2016). Genetic diversity at the Dhn3 locus in Turkish *Hordeum spontaneum* populations with comparative structural analyses. *Sci. Rep.* 6:20966. doi: 10.1038/srep20966
- Urao, T., Yamaguchi-Shinozaki, K., Urao, S., and Shinozaki, K. (1993). An Arabidopsis myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. *Plant Cell* 5, 1529–1539. doi: 10.1105/tpc.5.11.1529
- Wei, K., and Chen, H. (2018). Comparative functional genomics analysis of bHLH gene family in rice, maize and wheat. *BMC Plant Biol.* 18:309. doi: 10.1186/s12870-018-1529-5
- Wredenberg, A., Lagouge, M., Bratic, A., Metodiev, M. D., Spähr, H., Mourier, A., et al. (2013). MTERF3 regulates mitochondrial ribosome biogenesis in invertebrates and mammals. *PLoS Genet.* 9:e1003178. doi: 10.1371/journal.pgen.1003178
- Wright, S. (1951). The genetical structure of populations. *Ann. Eugen.* 15, 323–354. doi: 10.1111/j.1469-1809.1949.tb02451.x
- Xia, Y., Li, R., Bai, G., Siddique, K. H. M., Varshney, R. K., Baum, M., et al. (2017). Genetic variations of HvP5CS1 and their association with drought tolerance related traits in barley (*Hordeum vulgare* L.). *Sci. Rep.* 7:7870. doi: 10.1038/s41598-017-08393-0
- Xiong, H. B., Wang, J., Huang, C., Rochaix, J. D., Lin, F. M., Zhang, J. X., et al. (2020). mTERF8, a member of the mitochondrial transcription termination factor family, is involved in the transcription termination of chloroplast gene psbJ. *Plant Physiol.* 182, 408–423. doi: 10.1104/pp.19.00906
- Xu, D., Leister, D., and Kleine, T. (2017). Arabidopsis thaliana mTERF10 and mTERF11, but Not mTERF12, are involved in the response to salt stress. *Front. Plant Sci.* 8:1213. doi: 10.3389/fpls.2017.01213
- Yagi, Y., and Shiina, T. (2012). Evolutionary aspects of plastid proteins involved in transcription: the transcription of a tiny genome is mediated by a complicated machinery. *Transcription* 3, 290–294. doi: 10.4161/trns.21810
- Yakubovskaya, E., Guja, K. E., Mejia, E., Castano, S., Hambardjjeva, E., Choi, W. S., et al. (2012). Structure of the essential MTERF4:NSUN4 protein complex reveals how an MTERF protein collaborates to facilitate rRNA modification. *Structure* 20, 1940–1947. doi: 10.1016/j.str.2012.08.027
- Yang, Y. Z., Ding, S., Wang, Y., Wang, H. C., Liu, X. Y., Sun, F., et al. (2020). PPR20 is required for the cis-splicing of mitochondrial nad2 intron 3 and seed development in maize. *Plant Cell Physiol.* 61, 370–380. doi: 10.1093/pcp/pcz204
- Yin, X., Gao, Y., Song, S., Hassani, D., and Lu, J. (2021). Identification, characterization and functional analysis of grape (*Vitis vinifera* L.) mitochondrial transcription termination factor (mTERF) genes in responding to biotic stress and exogenous phytohormone. *BMC Genomics* 22:136. doi: 10.1186/s12864-021-07446-z
- Zhang, M., Kim, Y., Zong, J., Lin, H., Dievart, A., Li, H., et al. (2019). Genome-wide analysis of the barley non-specific lipid transfer protein gene family. *Crop J.* 7, 65–76. doi: 10.1016/j.cj.2018.07.009
- Zhang, Y., Cui, Y. L., Zhang, X. L., Yu, Q. B., Wang, X., Yuan, X. B., et al. (2018). A nuclear-encoded protein, mTERF6, mediates transcription termination of rpoA polycistron for plastid-encoded RNA polymerase-dependent chloroplast gene expression and chloroplast development. *Sci. Rep.* 8:11929. doi: 10.1038/s41598-018-30166-6
- Zhao, Y., Cai, M., Zhang, X., Li, Y., Zhang, J., Zhao, H., et al. (2014). Genome-wide identification, evolution and expression analysis of mTERF gene family in maize. *PLoS ONE* 9:e94126. doi: 10.1371/journal.pone.0094126

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Li, Pan, Yuan, Liu, Li, Wu, Wang and Cui. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Haplotype-Based Analysis of *Aegilops tauschii* Introgression Into Hard Red Winter Wheat and Its Impact on Productivity Traits

Moses Nyine¹, Elina Adhikari¹, Marshall Clinesmith², Robert Aiken², Bliss Betzen¹, Wei Wang¹, Dwight Davidson¹, Zitong Yu¹, Yuanwen Guo¹, Fei He¹, Alina Akhunova³, Katherine W. Jordan^{1,4}, Allan K. Fritz² and Eduard Akhunov^{1*}

¹ Department of Plant Pathology, Kansas State University, Manhattan, KS, United States, ² Department of Agronomy, Kansas State University, Manhattan, KS, United States, ³ Integrated Genomics Facility, Kansas State University, Manhattan, KS, United States, ⁴ United States Department of Agriculture, Agricultural Research Service Hard Winter Wheat Genetics Research Unit, Manhattan, KS, United States

OPEN ACCESS

Edited by:

Xianchun Xia,
Institute of Crop Sciences, Chinese
Academy of Agricultural
Sciences, China

Reviewed by:

Haiyang Liu,
Yangtze University, China
Sivakumar Chamarthi,
University of Arkansas, United States

*Correspondence:

Eduard Akhunov
eakhunov@ksu.edu

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 29 May 2021

Accepted: 20 July 2021

Published: 17 August 2021

Citation:

Nyine M, Adhikari E, Clinesmith M,
Aiken R, Betzen B, Wang W,
Davidson D, Yu Z, Guo Y, He F,
Akhunova A, Jordan KW, Fritz AK and
Akhunov E (2021) The
Haplotype-Based Analysis of *Aegilops*
tauschii Introgression Into Hard Red
Winter Wheat and Its Impact on
Productivity Traits.
Front. Plant Sci. 12:716955.
doi: 10.3389/fpls.2021.716955

The introgression from wild relatives have a great potential to broaden the availability of beneficial allelic diversity for crop improvement in breeding programs. Here, we assessed the impact of the introgression from 21 diverse accessions of *Aegilops tauschii*, the diploid ancestor of the wheat D genome, into 6 hard red winter wheat cultivars on yield and yield component traits. We used 5.2 million imputed D genome SNPs identified by the whole-genome sequencing of parental lines and the sequence-based genotyping of introgression population, including 351 BC₁F_{3:5} lines. Phenotyping data collected from the irrigated and non-irrigated field trials revealed that up to 23% of the introgression lines (ILs) produce more grain than the parents and check cultivars. Based on 16 yield stability statistics, the yield of 12 ILs (3.4%) was stable across treatments, years, and locations; 5 of these lines were also high yielding lines, producing 9.8% more grain than the average yield of check cultivars. The most significant SNP- and haplotype-trait associations were identified on chromosome arms 2DS and 6DL for the spikelet number per spike (SNS), on chromosome arms 2DS, 3DS, 5DS, and 7DS for grain length (GL) and on chromosome arms 1DL, 2DS, 6DL, and 7DS for grain width (GW). The introgression of haplotypes from *A. tauschii* parents was associated with an increase in SNS, which was positively correlated with a heading date (HD), whereas the haplotypes from hexaploid wheat parents were associated with an increase in GW. We show that the haplotypes on 2DS associated with an increase in the spikelet number and HD are linked with multiple introgressed alleles of *Ppd-D1* identified by the whole-genome sequencing of *A. tauschii* parents. Meanwhile, some introgressed haplotypes exhibited significant pleiotropic effects with the direction of effects on the yield component traits being largely consistent with the previously reported trade-offs, there were haplotype combinations associated with the positive trends in yield. The characterized repertoire

of the introgressed haplotypes derived from *A. tauschii* accessions with the combined positive effects on yield and yield component traits in elite germplasm provides a valuable source of alleles for improving the productivity of winter wheat by optimizing the contribution of component traits to yield.

Keywords: wheat, *Aegilops tauschii*, wild relative introgression, whole genome sequencing, haplotypes, genotype imputation, grain yield, yield component traits

INTRODUCTION

The gap between population expansion and food production is increasing due to marginal improvements in crop yield, which is attributed to a decline in soil fertility, pests and diseases, and climate change (Bailey-Serres et al., 2019). Wild relatives of wheat are a rich source of novel underutilized allelic diversity with a great potential to improve cultivated wheat through introgression (Placido et al., 2013; Zhang et al., 2017; Hao et al., 2020). The introgression from wild relatives into elite wheat cultivars was reported to increase pest and disease resistance (Periyannan et al., 2013; Saintenac et al., 2013), improve resilience toward environmental stress (Peleg et al., 2005; Placido et al., 2013), and increase yield (Pasquariello et al., 2020). The success of introgression breeding, however, could be affected by the negative epistasis between multiple alleles of wild and cultivated wheat (Nyine et al., 2020), especially in the low recombination regions of chromosomes, where the linkage with the negatively selected alleles could reduce the efficiency of selection for beneficial variants (Hill and Robertson, 1966).

Introgression could exhibit pleiotropic effects, affecting multiple, often unrelated traits not directly targeted by selection. For example, the introgression from *Aegilops ventricosa* into the chromosome 7D of wheat was associated with an increase in grain protein content and resistance to eyespot at the expense of reduced yield (Pasquariello et al., 2020). In durum wheat, the introgression of the *GN1-A1* gene from wild emmer increased grain weight by suppressing the fertility of distal florets, resulting in a negative correlation between grain number (GN) and grain weight (Golan et al., 2019). The introgression from *Agropyron elongatum* into the 7DL chromosome arm of wheat that is known to confer leaf rust resistance (*Lr19*) (Wang and Zhang, 1996) also influences root development, resulting in an improved adaption to water stress (Placido et al., 2013) and salinity (Dvorák et al., 1988), and increased biomass (BM) and yield (Reynolds et al., 2001).

Crop yield is a complex trait determined by many component traits, such as 1,000 grain weight (TGW), GN per spike, spikes per unit area, grain width (GW), area, length, etc. (Del Moral et al., 2003; Zhang et al., 2018; Du et al., 2020). Previous studies have shown that the introgression from wild and close relatives improve the yield of hexaploid wheat by changing yield component traits through the pleiotropic interaction between introgressed and background alleles of the hexaploid wheat (Jones et al., 2020). Significant trade-offs between yield, yield components, and yield stability have been reported in wheat. A study by Pennacchi et al. (2019) showed that yield and

yield stability have a negative linear relationship in most cases. Other factors such as HD, plant height (PH), and BM influence the source-sink ratio, which in turn affects the harvest index leading to a variation in yield and yield stability (Reynolds et al., 2001, 2020). Balancing the trade-off between yield components is therefore necessary to improve yield, maximize the yield potential, and improve the yield stability in wheat.

Sequence-based genotyping generates high-density SNP marker data that could be used to accurately detect the boundaries of genomic segments introgressed from wild relatives (Kuzay et al., 2019; Nyine et al., 2020), providing a unique opportunity to investigate the effect of introgression on trade-off between the traits affecting total yield. Even though whole-genome sequencing became less expensive, it is still not within the cost range that would allow wheat breeding programs to sequence large populations. Sequencing of founder lines at a high coverage depth and using these genotypes as a reference panel to impute missing and ungenotyped markers in the progeny characterized by low-coverage skim sequencing is a cost-effective alternative. The existing imputation algorithms (Browning and Browning, 2013; Swarts et al., 2014; Davies et al., 2016) provide a highly accurate whole-genome prediction of missing genotypes and were shown to increase the power of genome-wide association (GWAS) scans, thus enabling the identification of SNPs or haplotypes associated with the traits of interest (Li et al., 2010; Nyine et al., 2019). One of the advantages of the increased marker density provided by whole-genome sequencing is the ability to perform association mapping using haplotype information, which improves the detection of quantitative trait loci that would otherwise be missed when using single SNPs (Zhang et al., 2002; Lorenz et al., 2010).

Here, we investigated the impact of the introgression from *A. tauschii* into hard red winter wheat lines on yield and yield component traits and how haplotypes from *A. tauschii* at different genomic loci affect the component traits and total yield. For this purpose, we assessed the phenotypic variability of yield and the component traits, BM, and tenacious glume (Tg) traits in an introgression population derived from *A. tauschii* and hexaploid winter wheat phenotyped under irrigated and rainfed conditions. We applied the SNP diversity data generated by the whole-genome shotgun sequencing at 10× coverage level of the parental lines to impute genotypes in this population (Nyine et al., 2020). This strategy resulted in 5.2 million SNPs that enabled us to identify the introgressed *A. tauschii* haplotypes and assess their effects on the trait variation through the GWAS mapping and haplotype block effect analysis. This introgression population along with high-density SNP genotype data provides

a valuable resource for the effective deployment of *A. tauschii* haplotypes in winter wheat improvement programs.

MATERIALS AND METHODS

Plant Materials

The study population was described in detail by Nyine et al. (2020). In brief, the population was developed by crossing synthetic *A. tauschii*-wheat octoploid lines with recurrent hexaploid winter wheat lines. The octoploid lines were developed by crossing 5 hexaploid winter wheat lines with 21 diverse *A. tauschii* accessions. The resulting F₁ hybrid plants regenerated from the rescued embryos were treated with colchicine to produce synthetic octoploids (Dale et al., 2017). The octoploids were backcrossed once to the respective hexaploid wheat parents or to another wheat line. The BC₁F₁ plants (hexaploid) were self-pollinated and advanced by a single seed descent to the BC₁F₃ generation. Seeds from individual BC₁F₃ plants were bulked and grown in single rows in the field at the Kansas State University, Ashland Research Farm near Manhattan, KS, USA in 2016–2017 growing season. A total of 351 lines were selected from the entire population based on the ability to produce sufficient seeds to allow for yield testing, general fitness, threshability, and phenology corresponding to the adapted wheat cultivars.

Field Phenotyping

The population was phenotyped in 2018 and 2019 at Colby (Manhattan, KS, USA), and in 2020 at (Manhattan, KS, USA). In both locations and years, an augmented design was used to establish the trials. Plots were planted using a New Holland six-row wheat drill. Plot dimensions were 2.5 m long by 0.5 m wide and consisted of three rows with 18 cm row spacing. Starter fertilizer was applied with the seed at planting using granular 18-46-0 diammonium phosphate (DAP) at a rate of 168.1 kg/ha. Additional nitrogen was applied as a topdress in the spring using liquid 28-0-0 urea ammonium nitrate (UAN) at a rate of 67.3 kg/ha. A lateral irrigation system was used at Colby to ensure uniform germination and emergence as well as to provide additional water throughout the growing season in the irrigated treatment. Three hexaploid winter wheat lines well-adapted to Kansas environments (checks) and the hexaploid wheat parents were used as controls with at least three biological replications per block. In 2018 and 2019, two complete blocks were established and one block was irrigated (COI18 and COI19, respectively), whereas the other was rainfed/non-irrigated (CO18 and CO19, respectively), simulating optimal and farmer-field growth conditions. In 2020, only one block was grown under rainfed conditions (AS20).

The population was phenotyped for yield and yield components traits, BM traits, and Tg. Agrobase software (Mulitze, 1990) was used to adjust the grain yield (GY) (bushels per acre (BPA)), for spatial variability. The MARVIN seed imaging system [GTA Sensorik GmbH, Neubrandenburg, Germany) was used to assess the grain morphometric traits such as GN per sample, TGW, grain area (GA), GW, and grain length (GL) from the two technical replicates in 2018, and one measurement in 2019 and 2020. In 2019 and 2020, data were

collected for the number of spikes per square foot (SPSF) from two random points within a plot. The 1 × 1 ft square frame was dropped over two rows at least one foot away from the plot edges to avoid the border effect. In 2019, only one row within a frame was cut above the ground level for BM determination, whereas in 2020, both rows were sampled. Biomass samples were collected in paper bags and dried for at least 3 weeks at 32°C (90°F) before processing. We collected data on aboveground dry BM measured as the total weight of the dry sample without the bag, the number of spikes per sample (SPB), the average spikelet number per spike (SNS) from 10 random spikes, and grain weight after threshing [grain sample weight (GSW)]. During threshing, we scored samples for the presence and absence of the Tg trait depending on the threshability. Harvest index (HI) was calculated as the percentage of GSW relative to BM.

In 2020, data for HD were collected from each plot when approximately 50% of the spikes had emerged from the flag leaves. The number of days to heading were calculated as the difference between the heading and planting dates. After all the plots had completed heading, PH, in centimeters was measured on the same day from two random but representative main tillers per plot for the whole field. PH was measured as the distance from the ground surface to the first spikelet of the spike.

Genotyping

Whole-Genome Shotgun Sequencing of Parental Lines

Genomic DNA of 21 *A. tauschii* accessions and 6 hexaploid parents used to create the introgression population was extracted from the leaf tissues of 2-week-old seedlings grown in a greenhouse using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the protocol of the manufacturer. The quality and concentration of the DNA were assessed by using the PicoGreen dsDNA Assay Kit (Life Technologies, Carlsbad, CA, USA).

Genomic libraries for Illumina sequencing were constructed from ~2 µg of genomic DNA using the PCR-free Illumina protocol at the K-State Integrated Genomics Facility (IGF). The libraries were subjected to size selection using the Pippin Prep system (Sage Scientific, Beverly, MA, USA) to enrich for 400–600 bp fragments. The pooled barcoded libraries were sequenced using the NovaSeq instrument (2 × 150 bp run, S4 flow cell) at Kansas University Medical Center and NextSeq 500 (2 × 150 bp run) at IGF. The PCR-free whole-genome shotgun sequencing libraries generated from 27 parental lines ranged between 200 and 700 bp with an average of 450 bp (**Supplementary Figure 1**). Approximately 14 billion paired-end reads (150 bp long) were generated from the libraries with an average of 0.54 billion reads per sample (data are available at NCBI SRA database BioProject ID: PRJNA745927). The average number of reads per wheat line corresponded to ~10× genome coverage of parental lines. The raw reads with the Phred quality score <15, minimum length <50 bp, and adaptor sequences were filtered out using Trimmomatic v0.38-Java-1.8. The remaining filtered paired-end reads were mapped to the Chinese Spring (CS) RefSeq v1.0

(International Wheat Genome Sequencing Consortium, 2018) using the BWA-mem software v0.7.17. A total of 7.1 billion reads were aligned uniquely to the CS RefSeq v1.0.

The sam files generated by BWA-mem were converted to bam files using the SAMtools v1.10. The Picard Toolkit (<http://broadinstitute.github.io/picard>) was used to merge bam files from different lanes and sequencers into one bam file per sample. Reads that were aligned to multiple locations within the genome were identified and removed by SAMtools v1.10. The Picard Toolkit was used to prepare the merged unique aligned read bam files for GATK (McKenna et al., 2010) analysis. The preparatory steps included sorting, adding read groups, marking, and removing duplicate reads. The output deduplicated bam files were realigned around INDELs using GATK v3.7 and recalibrated with 90K SNPs (Wang et al., 2014) mapped to CS RefSeq v1.0 as the reference coordinates. The bam files were split into chromosome parts and indexed to reduce the memory and time required to process the files. GATK v3.7 was used to generate the genome variant call format (gvcf) files for each chromosome part. The gvcf files were split into 100 Mb chromosome windows and stored as genomicsDB using the genomicsDBImport tool in GATK v4.0. Joint genotyping of variants from each database was done using GATK v4.0 HaplotypeCaller. The flag `<-allow-old-rms-mapping-quality-annotation-data>` was set to enable the processing of gvcf files generated by GATK v3.7. All vcf files corresponding to the A, B, and D genomes were concatenated with concat, a Perl-based utility in vcftools v0.1.13. A custom Perl script was used to convert the CS RefSeq v1.0 part coordinates in the concatenated vcf to full coordinates after which vcf-filter tools v0.1.13 were used to remove INDELs, multi-allelic loci, sites with missing data and minor allele frequency (MAF) below 0.05. The filtered SNPs were phased using the Beagle software v4.1 (Browning and Browning, 2013).

The GATK v4.0 HaplotypeCaller identified ~99 million variants including SNPs and INDELs from reads uniquely aligned to the D genome of CS RefSeq v1.0. After excluding INDELs, multi-allelic loci, sites with missing data, and MAF <0.05, 20 million SNPs were retained. These were used to impute missing and ungenotyped SNPs in the D genome of the introgression population.

Genotype Imputation

Sequence-based genotyping of the introgression population was performed previously by Nyine et al. (2020). SNPs with MAF <0.01 were excluded from the raw vcf file using vcf-filter tools v0.1.13. The program conform-gt (<https://faculty.washington.edu/browning/conform-gt.html>) was used to check the concordance of D genome SNP positions between the introgression population and the SNPs from the parental lines genotyped by whole-genome shotgun sequencing. Missing and ungenotyped SNPs in the D genome of the introgression population were imputed from the parental lines using Beagle v5.0. A custom Perl script was used to filter out the imputed SNPs with genotype probability below 0.7 and MAF <0.05, which resulted in 5.2 million SNPs. The filtered SNPs were used in the downstream analyses such as GWAS mapping and identification of the introgressed haplotype blocks.

Introgressed Haplotype Detection

Genetic divergence between the parental lines affects the probability of an accurate detection of the introgressed segments from wild relatives. We used two introgression families, one created by crossing hexaploid wheat with *A. tauschii* ssp. *strangulata* (KanMark x TA1642, aka FAM93) and another one created by crossing hexaploid wheat with *A. tauschii* ssp. *tauschii* (Danby x TA2388, aka FAM97) to identify the introgressed haplotype blocks. FAM93 had 21 introgression lines (ILs), whereas FAM97 had 23 ILs. The R package HaploBlocker (Pook et al., 2019) was used to infer haplotype blocks from 5.2 million SNP sites. Recombinant inbred lines from each family were analyzed together with 21 *A. tauschii* and 6 hexaploid parents. The HaploBlocker parameters used in block calculation were `node_min = 2` (default is 5), `overlap_remove = TRUE`, `bp_map`, and `equal_remove = TRUE`. The parameter `node_min` was used to control the number of haplotypes per node during a cluster-merging step of the block calculation function of HaploBlocker. The reduction in `node_min` was necessary to account for the low number of haplotype variants within these families. To maintain the SNP position in the haplotype block library, a vector of SNPs was provided *via* the parameter `bp_map` and prior to block calculation, SNPs in perfect linkage disequilibrium were removed by setting the parameter `equal_remove = TRUE`. Overlapping haplotypes were removed by setting parameter `overlap_remove = TRUE`. Custom R and Perl scripts were used to calculate the haplotype block length using the information from haplotype block start and end coordinates. All monomorphic haplotypes between the two parental lines were excluded from a haplotype matrix before calculating the frequency of the introgressed haplotypes per chromosome.

Phenotype Data Analysis

Trait Stability and Heritability

Yield stability is an important trait, which reflects the productivity of the crop under various growth conditions. No single statistic, however, is accurate enough to predict it due to a high variability of genotype and genotype by environment interaction effects. In this study, we used 16 different statistics, including parametric [such as mean variance component (θ_i), GE variance component [$\theta_{(i)}$], Wricke's ecovalence (W_i^2), regression coefficient ($*b_i$), deviation from regression (S_{di}^2), Shukla's stability variance (σ_i^2), coefficient of variance (CV $_i$), and Kang's rank-sum (Kang or KR)] and non-parametric [such as Huhn's and Nassar and Huhn's ($S^{(1)}$, $S^{(2)}$, $S^{(3)}$, $S^{(4)}$, $S^{(5)}$, and $S^{(6)}$), and the methods of Thennarasu ($NP^{(1)}$, $NP^{(2)}$, $NP^{(3)}$, and $NP^{(4)}$)] to rank the ILs for yield stability based on their performance across years, locations, and treatments. The description and properties of the statistics are documented at the website: <https://manzik.com/stabilitysoft/>. The analysis was implemented in R using a script from Pour-Aboughadareh et al. (2019), which is available at: <https://github.com/pour-aboughadareh/stabilitysoft>. The most stable and/or high yielding lines were identified by sorting them according to their rankings.

Broad sense heritability (H^2) and best linear unbiased predictions (BLUPs) for yield and the component traits were calculated from the 2018 and 2019 data. A mixed linear model with restricted maximum likelihood implemented in R package

lme4 was used to generate the variance components (var) that were used to calculate H^2 as follows:

Model = lmer(Trait ~ Trt
+ (1|Genotype) + (1|Genotype:Trt), data = trait_data)

$H^2 = \text{var}(\text{Genotype}) / [\text{var}(\text{Genotype}) + \text{var}(\text{Genotype:Trt}) / \text{No. of Trt} + \text{var}(\text{Residual}) / \text{No. of Trt}]$ where Trt refers to irrigated and non-irrigated treatments. BLUPs were extracted from the linear mixed model as the random effects of genotypes.

Multiple comparisons for the effect of treatment on yield and yield components traits in the introgression population relative to the controls were performed using least squares (LS) means with “Tukey” adjustment method and $\alpha = 0.05$. To further assess the impact of introgressed haplotypes on the traits, lines were sorted in a descending order for each treatment and location/year. The percentage of the ILs that performed better than the best parental lines and checks [percentage of top-performing lines (PTPL)] was calculated for each trait. Similarly, ILs that produced more grains than both parents and checks were identified based on the mean spatial adjusted yield. The percentage increase in yield was calculated as follows:

$$\Delta \bar{y}_{ij} = 100 \times \frac{\bar{y}_i - \bar{y}_j}{\bar{y}_j}$$

where $\Delta \bar{y}_{ij}$ is the percentage change in mean yield, $\bar{y}_i$ is the mean yield for the high yielding ILs, and $\bar{y}_j$ is the mean yield for the controls (parents and checks).

Trade-Off Between Traits

The relationship between yield, yield components, and BM traits were assessed by calculating the Pearson's correlation coefficients. We compared the trend of correlations from different treatments (irrigated/non-irrigated) and years to determine the extent of trade-off between traits within the introgression population and how the environment influenced them.

GWAS Mapping

The genome-wide association analysis was used to identify the genomic regions with SNPs and haplotypes that have a significant association with the traits. We tested the association of 5.2 million SNPs from the D genome with the traits phenotyped in different treatments and years. A total of 15,967 haplotype block windows were identified from 5.2 million SNPs using the R package HaploBlocker v1.5.2 (Pook et al., 2019). Default parameters for HaploBlocker were used except node_min, which were reduced to two (default is five) as most genomic intervals in our data set had <5 haplotype variants. Overlapping haplotypes were removed using the parameter overlap_remove = TRUE, and the SNP coordinates were included in the haplotype library via the parameter bp_map. Haplotype blocks were split into windows by setting the parameter return_dataset = TRUE in the block_windowdataset() function. The haplotype variants within a given chromosome interval were recorded as 0, 1, 2, or 3 depending on the total number of haplotype variants present within the interval. In both cases, a mixed linear

model implemented in the R package GAPIT was used to detect the associations. To control for a population structure in SNP-based analyses, 100,000 randomly selected markers were used to calculate the principal components. In haplotype-block-based GWAS, all haplotype blocks were used to calculate the principal components. In both cases, only the first three principal components were used to control the population structure. Two threshold options were used to identify significant associations including a more stringent Bonferroni correction and a less stringent Benjamini and Hochberg (1995) false discovery rate (FDR) at 5% significance level. To control for the effect of treatment and year, GWAS based on BLUPs was also performed, and significant associations were reported only when there were SNP- or haplotype-trait association in at least two independent trials or in the BLUP-based analysis.

SNP- and Haplotype-Trait Effects

Haplotype variation at loci with significant SNPs and their effects on traits in the introgression population were analyzed. The R package HaploBlocker v1.5.2 (Pook et al., 2019) was used to infer haplotypes at the genomic loci with significant SNP-trait associations. Heatmap.2 function provided in the R package gplots was used to visualize the variation in haplotypes from hexaploid wheat and *A. tauschii*. However, at the *Ppd-D1* locus, a visual comparison of the SNPs from the parental lines was done, and the SNPs were annotated using the snpEff v4.3 software to resolve haplotype variants in the *A. tauschii* lines that could not be clearly distinguished by HaploBlocker. SNPs significant at $\text{FDR} \leq 0.05$ and their estimated allelic effects were selected from the association mapping results and used to verify if the haplotype effect corresponded to the observed phenotype in the introgression population. The mean and the SD of the phenotype were calculated for each group of lines carrying a similar haplotype, and the difference between the means was compared using Tukey's honestly significant difference and the Student's *t*-test.

RESULTS

Trait Variation in the Introgression Population

Broad sense H^2 of GY was 0.7 in 2018 and 0.64 in 2019, whereas for the yield component traits such as TGW, GA, GW, and GL, H^2 values were 0.85, 0.89, 0.83, and 0.95, respectively. The agronomic performance of the ILs relative to the controls (parent/checks) was assessed by comparing their yield and yield component traits under different treatments. The effect of treatment on yield was significant in 2018 ($p < 2.2\text{e-}16$), but not in 2019 ($p = 0.24$) at 95% confidence level. The latter is partially associated with more abundant rainfall in 2019 that reduced the difference in the water availability stress levels between the irrigated and non-irrigated field trials in Colby, KS, USA. Based on the LS means, significant differences in yield between controls and ILs was observed in 2019 and 2020, but not in 2018 (Table 1). The yield data collected from irrigated and rainfed (non-irrigated) field trials conducted between 2018 and 2020 revealed that up to 23% of the lines with introgressions produce

TABLE 1 | Comparison of grain yield (GY) between introgression lines (ILs; progeny) and controls (checks/hexaploid parents) per treatment within a year using least squares (LS) means.

Year	Treatment	Group	LSmean	SE	df	Lower.CL	Upper.CL	Group
2018	Irrigated	Progeny	68.6	0.77	410	66.8	70.5	a
2018	Irrigated	Check	72.1	7.64	410	53.8	90.4	a
2018	Irrigated	Parent	75.7	3.6	410	67.1	84.3	a
2018	Non-Irrigated	Progeny	50.1	6.43	395	34.6	65.5	a
2018	Non-Irrigated	Check	51.6	0.66	395	50	53.2	a
2018	Non-Irrigated	Parent	57.1	2.95	395	50	64.2	a
2019	Irrigated	Progeny	103	0.82	379	100.6	105	a
2019	Irrigated	Parent	119	3.8	379	109.5	128	b
2019	Irrigated	Check	121	9.05	379	99.3	143	ab
2019	Non-Irrigated	Progeny	101	0.86	382	99.2	103	a
2019	Non-Irrigated	Parent	116	3.85	382	106.3	125	b
2019	Non-Irrigated	Check	116	9.44	382	93.5	139	ab
2020	Non-Irrigated	Progeny	47.1	0.43	385	46	48.1	a
2020	Non-Irrigated	Parent	55.3	2.01	385	50.5	60.1	b
2020	Non-Irrigated	Check	59.3	4.79	385	47.8	70.8	b

The 2018 and 2019 trials were conducted at Colby, whereas the 2020 trial was conducted at Ashland, Manhattan, KS, USA. Groups with similar letters are not significantly different at 95% confidence level.

more grain than the controls (**Supplementary Figure 2**). In 2018, however, 3.2 and 48% of the ILs produced more grain than the parental lines and checks, respectively, in the non-irrigated trial (CO18), suggesting that the check cultivars were more sensitive to drought stress than the parental lines.

The proportion of ILs outperforming the checks and parental lines for the measured traits varied between treatments and years with a minimum of 0.8% for BM in the non-irrigated trial in 2019 (CO19) and a maximum of 73% for TGW in the 2018 irrigated trial (COI18). The percentage increase in yield for the ILs that outperformed both checks and parent varied from 11 to 57%, whereas the number of lines that produced more grains varied from 6 to 94 (**Table 2**). Under drought stress conditions in 2018 (CO18), the mean yield of top-performing ILs was 1.6- and 1.4-folds higher than the checks and parents, respectively (**Figure 1**). These results suggest that some ILs carry the alleles that confer drought tolerance, thus ensuring high productivity under stressful conditions. The highest yield potential of both ILs and controls was observed in 2019. The mean of the top yielding ILs reached 134 BPA, whereas that of the parental lines and checks reached 119 and 121 BPA, respectively (**Table 2**).

The yield stability analyses were performed to identify ILs that are both high yielding and stable under various environmental conditions. Ranking by mean yield showed that 6% of the lines carrying introgression produced more grains than most parental lines, except Larry (**Supplementary File 1**). The yield of these lines ranged from 84.4 to 92.7 BPA. The average rank (AR) of 16 stability statistics revealed 12 lines with introgressions showing high yield stability. Five of these lines were both stable and high yielding according to KR when compared to the controls. The yield of these five ILs (KS15SGDCB110-11, KS15SGDCB098-1, KS15SGDCB103-6, KS15SGDCB098-14, and KS15SGDCB111-1) varied between 82 and 93 BPA. The yield of the most stable and high yielding IL (92.7 BPA) was 9.8% higher than the average

TABLE 2 | Percentage of mean yield difference between top-performing ILs and controls (parents and checks) per treatment in each year and location.

Group ID	No. of IL	Mean yield IL	Mean yield controls	% yield diff.
AS20_ILP	6	68.6	55.2	24.3
AS20_ILC	6	68.6	59.3	15.8
CO18_ILP	21	78.7	57.1	37.8
CO18_ILC	21	78.7	50.1	57.1
COI18_ILP	94	86.9	75.7	14.8
COI18_ILC	94	86.9	72.1	20.6
CO19_ILP	11	130.5	115.5	13.0
CO19_ILC	11	130.5	116.1	12.4
COI19_ILP	6	134.2	119.2	12.6
COI19_ILC	6	134.2	121.0	10.9

AS, Ashland non-irrigated trial; CO, Colby non-irrigated trial; COI, Colby irrigated trial; ILP, introgression lines by parents' GY comparison; ILC, introgression lines by checks GY comparison.

yield of the controls (84.4 BPA). These results indicate that novel alleles from *A. tauschii* have the potential to increase the adaptive potential of hard red winter wheat to different environmental conditions. In addition, the stability statistics could help to prioritize ILs for deployment in different agroecological zones depending on their ranking in stability and yield. Lines that are moderately high yielding but show good yield stability could be deployed in marginal environments, whereas less stable but high yielding lines could be deployed in less stressful environments to achieve high productivity.

Harvest index, a measure of source-sink capacity, was also assessed for stability in irrigated and rainfed trials. About 92 ILs showed a higher average HI (47.4–52.8) than the best parental line KS061406LN-26 (47.3). The AR based on the 16 stability

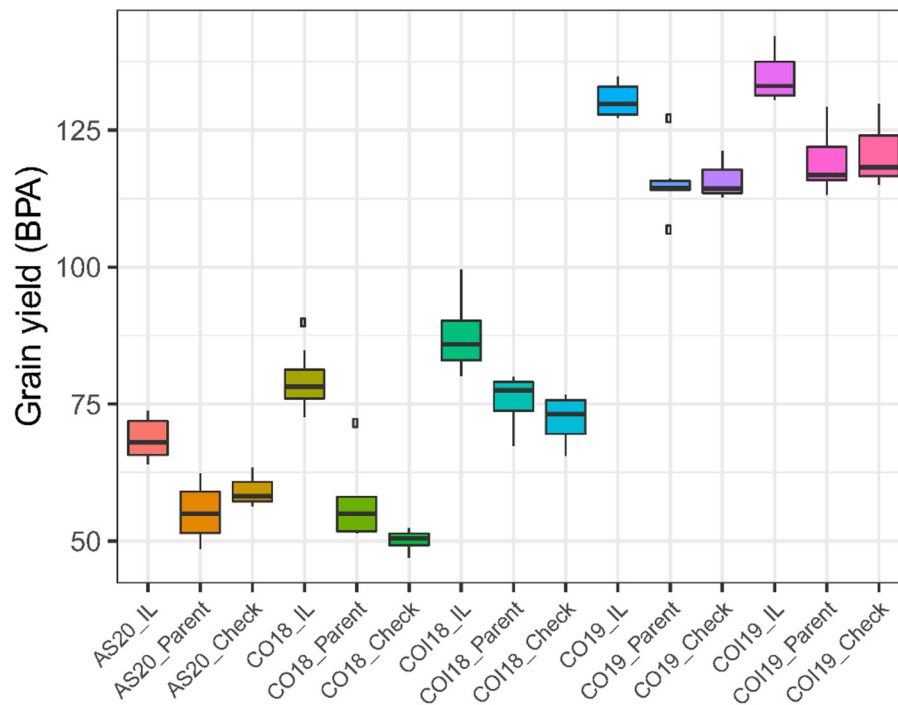


FIGURE 1 | Boxplots comparing the mean grain yield (GY) between the top-performing introgression lines (ILs) and the controls (parents and checks) across treatments, years, and locations. AS20 refers to Ashland rainfed trial in 2020, CO18 is Colby rainfed trial in 2018, COI18 is Colby irrigated trial in 2018, CO19 is Colby rainfed trial in 2019, and COI19 is Colby irrigated trial in 2019.

statistics placed 11 out of 92 lines in the top 20 most stable lines (**Supplementary File 2**). Line KS15SGDCB111-1, which is high yielding and stable, also ranked in the top five lines with a stable and high average HI.

Trade-Off Between Yield and Yield Component Traits

Pearson's correlation coefficients between the average yield and yield stability based on AR of the 16 stability statistics was -0.44 ($p < 0.001$; **Supplementary File 1**). However, the correlation between yield and KR was -0.71 ($p < 0.001$), indicating that the most stable ILs were not necessarily the highest grain yielders although there were some exceptions. Similarly, the correlation between average HI and AR was -0.42 ($p < 2.2e-16$), whereas the correlation between HI and KR was -0.73 ($p < 2.2e-16$; **Supplementary File 2**).

The trade-off between yield and yield components was influenced by the treatment, year, and location as evidenced by a variation in the levels of correlations (**Supplementary File 3**). Higher positive correlations were observed among grain morphometric traits such as TGW, GA, GW, and GL, ranging from 0.13 (between GW and GL) to 0.96 (between TGW and GA) (**Figure 2**). HI and GSW positively correlated with GY, whereas the correlation between GY and GN was positive but nonsignificant in all trials, except for the Colby irrigated trial in 2019 (COI19) (**Supplementary File 3**). BM correlated negatively with HI but showed a positive correlation with GSW

(**Supplementary File 3**). In some cases, an increase in the SNS resulted in a reduction in the TGW, GA, GW, or GL, which was consistent with the previously observed trade-off between these traits (Kuzay et al., 2019). In contrast, HD positively correlated with the SNS and PH, which is in agreement with the previous findings (Shaw et al., 2013; Muqaddasi et al., 2019).

To further understand the contribution of different yield components to the final yield, we compared the phenotypes of the top yielding ILs to those of the controls across all treatments (**Supplementary File 4**). In the CO18 trial under non-irrigated conditions, the ILs that outperformed the controls in yield had the highest TGW, GA, and GL, whereas under irrigated conditions (COI18), all yield component traits showed the highest levels of expression in the top yielding ILs. The top yielding ILs in the CO19 trial had the highest HI, GW, SNS, and BM while TGW and GA were comparable to those of the parental lines. In the COI19 trial, the TGW, GA, and GW traits contributed more toward the final yield compared to the GL, HI, and BM traits. In the AS20 trial, high levels of heterogeneity were observed among the top yielding lines for the TGW, GA, GW, and GL traits. However, these lines showed a higher BM than the controls, resulting in a reduced HI.

Previously, it was suggested that the introgression from wild relatives might have a negative impact on agronomic traits due to a negative epistasis between the alleles of wild and cultivated wheat (Nyine et al., 2020). We investigated the relationship between the total size of the introgressed

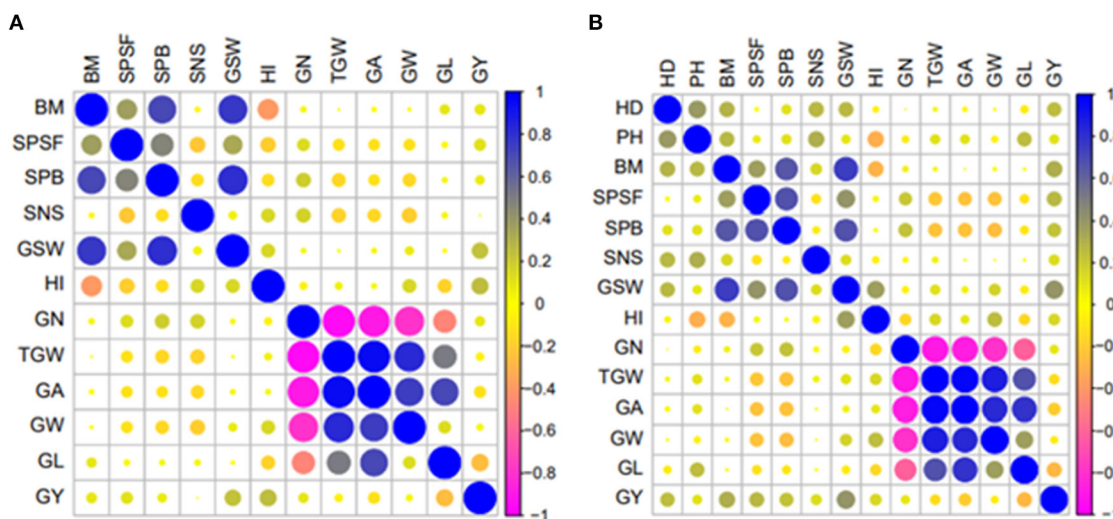


FIGURE 2 | Pearson's correlation coefficients between yield and yield component traits at Colby in 2018 rainfed trial (A) and Ashland in 2020 rainfed trial (B). Where HD, heading date; PH, plant height; BM, aboveground dry biomass; SPSF, spikes per square foot; SPB, spikes per bag; SNS, spikelet number per spike; GSW, grain sample weight; HI, harvest index; GN, grain number; TGW, 1,000 grain weight; GA, grain area; GW, grain width; GL, grain length; GY, grain yield.

genomic segments and phenotype. We found a positive linear relationship between GA, GL, SNS, and the total size of the introgressed segments (Figure 3, Supplementary File 5). For the TGW, however, a positive linear relationship was only observed under drought stress conditions indicating that some wheat lines with large introgressions are efficient in utilizing the limited soil moisture and nutrients during grain filling. There was a negative relationship between GY, HI, GW, TGW under irrigated conditions, and the size of introgression.

Genotyping and Imputation

To identify *A. tauschii* haplotypes in the D genome of introgression population, we generated high-density SNP data. By the whole-genome sequencing of 6 hexaploid parental lines and 21 *A. tauschii* accessions used for generating octoploid parents, we identified about 20 million high-quality SNP variants ($MAF \geq 0.05$) and used them for genotype imputation in the introgression population genotyped by complexity-reduced sequencing. The total number of D genome SNPs retained after filtering out SNPs with genotype probability below 0.7 and $MAF < 0.05$ was 5.2 million.

Haplotypic Variation Between *ssp. strangulata* and *ssp. tauschii* Families

Using HaploBlocker v1.5.2, we identified 4,764 and 6,429 non-overlapping haplotype blocks in the *A. tauschii* *ssp. strangulata* (FAM93) and *A. tauschii* *ssp. tauschii* (FAM97) families, respectively. After filtering out the monomorphic haplotypes between the parental lines, 869 (18%) and 3,020 (47%) segregating haplotypes were retained in FAM93 and FAM97, respectively (Table 3, Supplementary File 6). The low proportion of segregating haplotypes between hexaploid wheat and *ssp. strangulata* D genomes is in agreement with the finding that *A. tauschii* *ssp. strangulata* was the donor of the D genome

of hexaploid wheat (Wang et al., 2013). These results also suggest that the high similarity between the genome of *ssp. strangulata* and the D genome of hexaploid wheat could result in the underestimation of the proportion of the introgressed haplotypes. The average genome-wide haplotype block length in FAM93 was higher (2 Mb) than that in the FAM97 (1 Mb) (Supplementary File 6). There was a significant difference in the introgressed haplotype length between the lines in FAM93 and FAM97 based on the *t*-test ($p = 3.1e-16$). The longest haplotype introgressed in all lines from FAM93 was 44 Mb on chromosome arm 3DL, whereas in FAM97 only four lines had a haplotype >32 Mb on chromosome arm 7DL. The number of segregating haplotypes in FAM93 varies from 32 (3D) to 336 (2D), whereas in FAM97 the number of segregating haplotypes varies from 173 (3D) to 617 (5D) (Table 3). In FAM93 and FAM97, the average frequency of each haplotype from *A. tauschii* parents in the ILs was 11 and 4, respectively (Supplementary File 6).

SNP- and Haplotype-Based GWAS Mapping

Genome-wide association study was performed in the *A. tauschii* introgression population to assess the effects of introgression into the D genome on the variance of traits related to BM, yield and yield components, and Tg. The marker-trait association analyses were based on both individual SNPs and haplotype blocks identified by HaploBlocker from the 5.2 million imputed variants. We report only those associations that are replicated in at least two independent field trials and show a significant association with both SNPs and haplotypes at FDR 0.05 (Supplementary Table 1). Several genomic loci with significant associations distributed on the D genome chromosomes were detected for GL, GW, and SNS. For other traits such as GY, TGW, GN, GA, HI, BM, GSW, and SPSE, no consistent associations replicated in independent trials were detected.

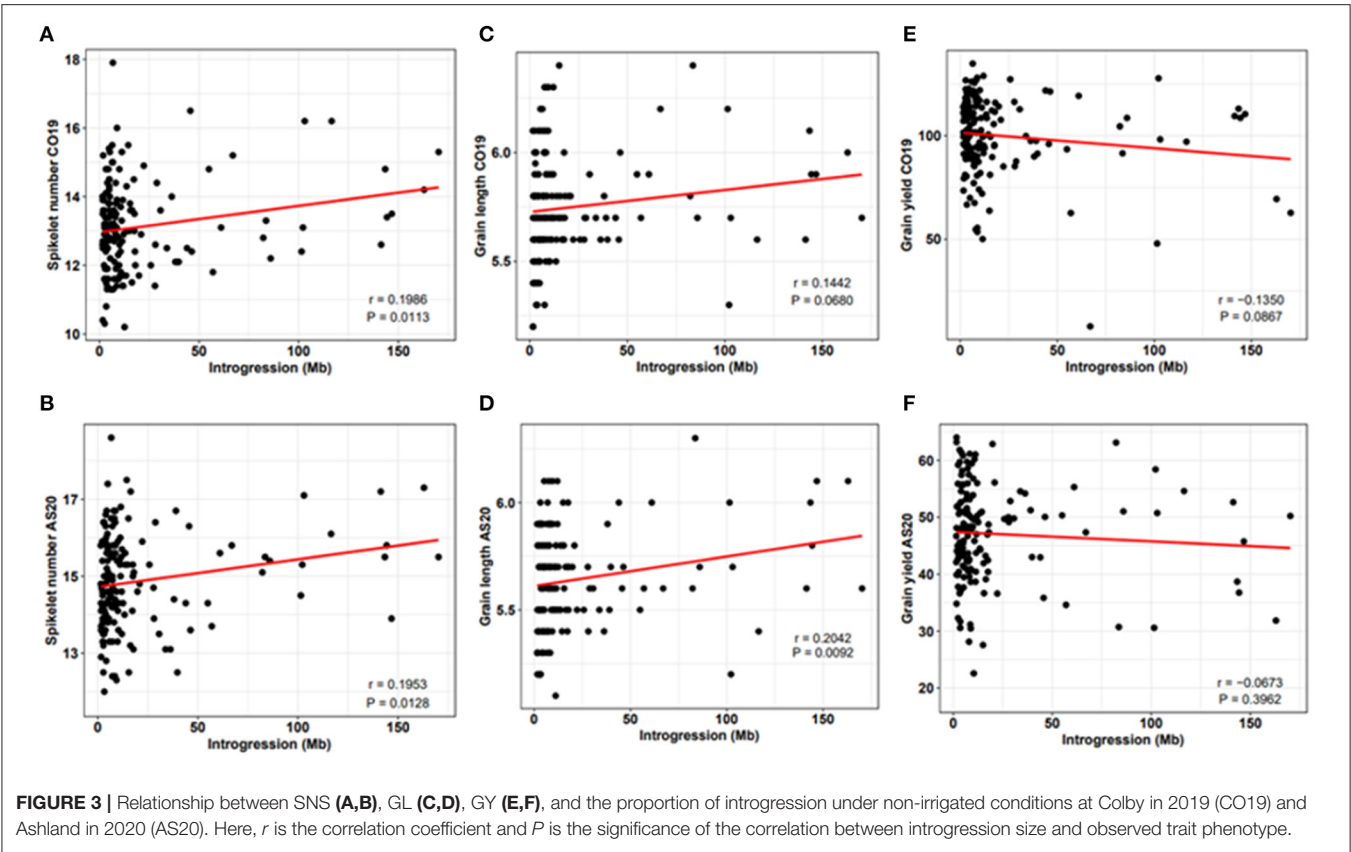


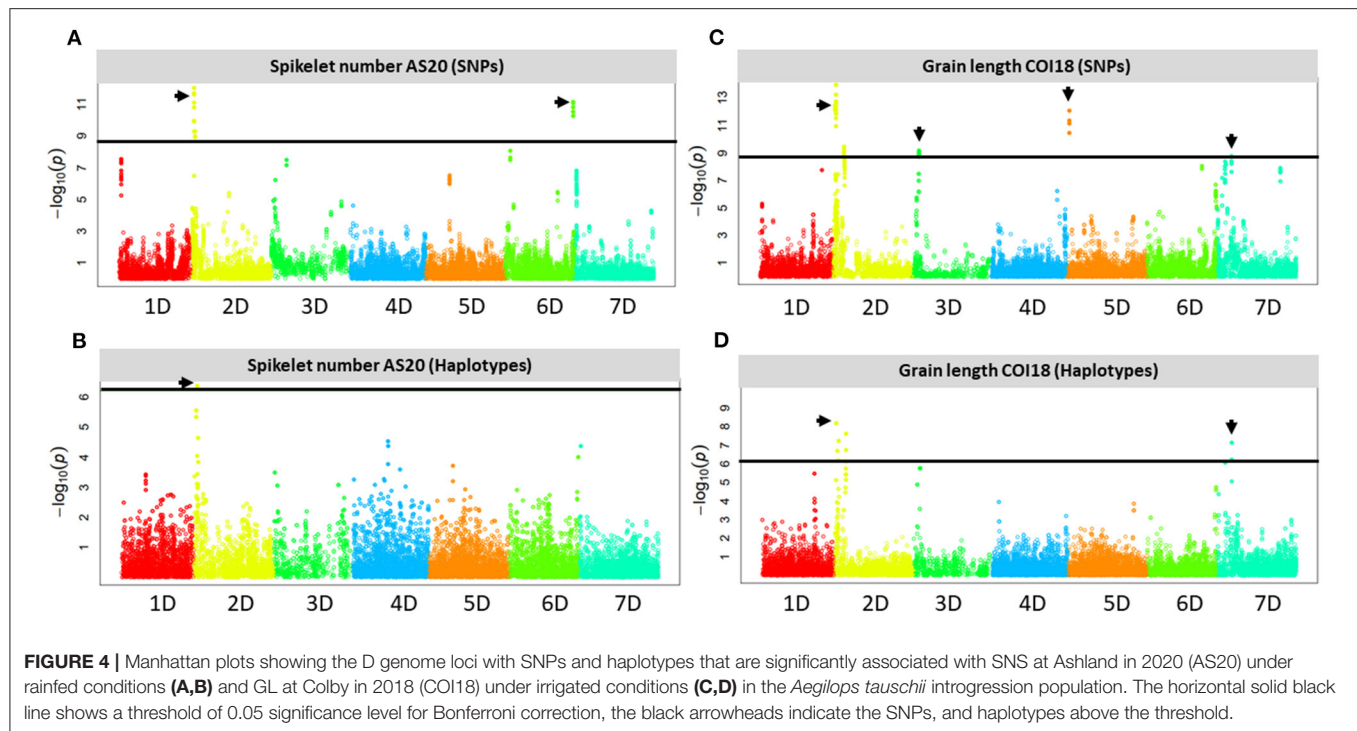
TABLE 3 | Variation of introgressed haplotypes between *Aegilops tauschii* ssp. *strangulata* (FAM93) and *A. tauschii* ssp. *tauschii* (FAM97) families.

Family	Pedigree	Subspecies	Chrom	NH	MHL (bp)	MHF	Range (bp)
FAM93	KanMark × TA1642	strangulata	chr1D	105	1,896,815	4.6	1,788–8,639,976
FAM93	KanMark × TA1642	strangulata	chr2D	336	1,326,620	20.0	813–30,656,640
FAM93	KanMark × TA1642	strangulata	chr3D	32	6,653,571	4.9	2,525–44,264,371
FAM93	KanMark × TA1642	strangulata	chr4D	143	1,865,609	4.3	2,942–12,070,089
FAM93	KanMark × TA1642	strangulata	chr5D	138	2,302,767	7.1	556–15,103,840
FAM93	KanMark × TA1642	strangulata	chr6D	65	2,519,945	3.0	3,455–13,483,300
FAM93	KanMark × TA1642	strangulata	chr7D	50	2,667,416	2.04	1,627–11,349,447
FAM97	Danby × TA2388	tauschii	chr1D	495	829,609	3.1	762–9,169,049
FAM97	Danby × TA2388	tauschii	chr2D	352	1,305,439	1.8	813–15,991,790
FAM97	Danby × TA2388	tauschii	chr3D	173	2,968,356	3.0	884–27,225,468
FAM97	Danby × TA2388	tauschii	chr4D	540	750,352	0.8	1,982–20,010,593
FAM97	Danby × TA2388	tauschii	chr5D	617	770,483	10.1	1,243–15,396,133
FAM97	Danby × TA2388	tauschii	chr6D	373	830,168	3.8	764–23,208,199
FAM97	Danby × TA2388	tauschii	chr7D	465	1,000,899	3.7	1,397–32,547,936

NH, Number of haplotype blocks; MHL, Mean haplotype block length; MHF, Mean haplotype frequency.

We identified multiple significant SNP- and haplotype-trait associations from all trials on chromosome arms 2DS and 7DS for GL (Figure 4). The most significant SNPs were located in haplotype block windows 22,262,355–22,289,017, 30,582,113–30,595,115, and 80,864,297–81,398,316 bp on chromosome arm 2DS, and 11,024,311–11,374,767 bp on chromosome arm 7DS (Supplementary Table 1). Association analysis based on BLUPs confirmed the results obtained

from individual trials for GL on these two chromosomes. Other significant associations detected in at least two trials were identified on chromosome arms 3DS and 5DS (Supplementary File 7). We identified haplotypes with significant SNPs associated with GW on 1DL, 2DS, 6DL, and 7DS from at least two independent trials that were confirmed by BLUP-based analysis (Supplementary File 7). Haplotype block windows



65,964,778–66,124,103 and 66,265,325–66,266,089 bp showed the most significant association on 2DS.

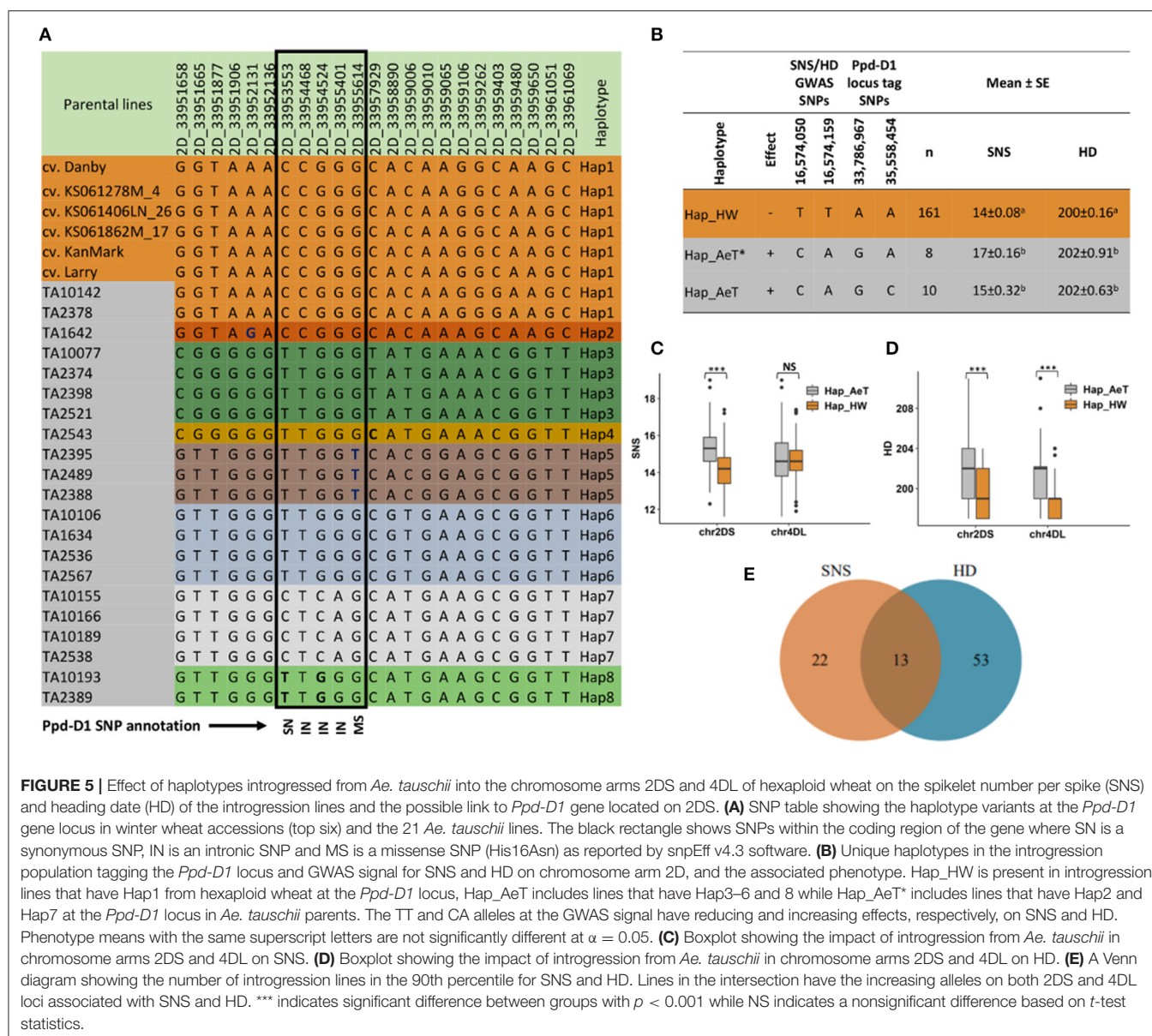
At 95% confidence level, the most significant SNP-trait associations were identified on chromosome arms 2DS and 6DL for SNS from the three independent trials (COI19, COI18, and AS20), (**Figure 4**, **Supplementary File 7**). The most significant associations are located at 16.5 and 463.8 Mb on 2DS and 6DL, respectively. Haplotype-trait analysis confirmed the association on 2DS for SNS at the 16.5 Mb locus located within the haplotype block window 16,497,666–16,548,006 bp. At $FDR < 0.05$, there was no haplotype block window on 6DL locus that overlapped with a significant SNP-trait association.

Previous studies have shown that SNS is linked to HD (Shaw et al., 2013; Muqaddasi et al., 2019). In the current study, we detected significant associations with SNS on chromosome arms 2DS and 6DL. We had 1 year data for HD and PH collected from Ashland in 2020, which provided us a good opportunity to validate this link in the *A. tauschii* introgression population. GWAS mapping detected significant associations with HD on chromosome arms 2DS and 4DL, whereas all D genome chromosomes showed a significant association with PH but the strongest signals were observed on 1DS, 3DS, and 6DL. The haplotype block window 16,548,753–16,639,561 bp on 2DS with the most significant SNPs for HD overlapped the locus showing a significant association with SNS, which is in proximity to another haplotype block overlapping with the most significant SNPs for SNS (16,497,666–16,548,006 bp). These results suggest that the expression of these two traits could be co-regulated.

For HD, the haplotype block windows on chromosome arm 4DL 442,735,095–442,751,954 bp and 459,271,685–459,290,731

bp had the most significant SNP-trait associations. The three traits (SNS, HD, and PH) are known to be affected by the *Rht8* and *Ppd-D1* genes on 2DS, in addition to *Rht1* on 4D, which control PH and flowering time (Borojevic and Borojevic, 2005; Chen et al., 2018). Due to the lack of SNPs located near the *Ppd-D1* gene locus at ~34 Mb (33,961,438–33,951,651 bp interval in CS RefSeq v1.0), we could not directly validate its association with these traits. However, significant associations for SNS were detected at ~3 Mb next to the *Ppd-D1* locus in the COI19 and AS20 trials on haplotype blocks 2D: 30,192,335–30,264,745 bp and 2D: 28,829,778–28,937,705 bp, respectively. In the parental lines with high-density SNPs (~20 million), the *Ppd-D1* locus had SNPs, which allowed us to precisely map the haplotypes from *A. tauschii* and hexaploid wheat lines. The results obtained from HaploBlocker showed that all hexaploid parents carry an identical haplotype, which is distinct from that of *A. tauschii* accessions.

By using SNPs identified by the whole-genome sequencing of parental lines, we characterized a haplotypic diversity at the *Ppd-D1* locus (**Figure 5A**). All hexaploid wheat lines carried the same *Ppd-D1* haplotype (Hap1) while seven haplotypes of the *Ppd-D1* gene (Hap2–Hap8) were identified in *A. tauschii*. The whole-genome sequencing of 21 *A. tauschii* revealed a broader range of *Ppd-D1* diversity compared to a previous study (Guo et al., 2009), which identified only three *Ppd-D1* haplotypes. The *A. tauschii* ssp. *strangulata* accessions carried the haplotypes that were identical to hexaploid wheat, except for Hap2 in TA1642, which had one SNP at position 33,952,131 bp (**Figure 5A**). The *Ppd-D1* genic region in *A. tauschii* ssp. *tauschii* accessions has one synonymous (SN), three intronic (IN), and one missense (MS) SNPs. The MS variant at position 33,955,614 bp results in His16Asn change, which is predicted to have a moderate



functional impact, and only present in the lines with haplotype Hap5 (Figure 5A). Next, we inferred the parental haplotypes of the *Ppd-D1* locus in the introgression population by using SNPs within the ~1–2 Mb region surrounding the *Ppd-D1* locus. About 82% of the ILs carried haplotype Hap1.

Further, we evaluated the linkage of *Ppd-D1* haplotypes with SNP alleles showing a significant association with a variation in SNS and HD. For this purpose, we used two SNP sites, 2D_33786967 and 2D_35558454, which flank the *Ppd-D1* locus on both sides and have genotyping information in the introgression population. We compared them to SNP alleles that were significantly associated with SNS and HD in a haplotype block window 2D: 16,548,753–16,639,561 bp (2D_16574050 and 2D_16574159), spanning ~17 Mb region (Figure 5B). We found that the GWAS alleles associated with an increase in SNS and HD in the introgression population are also linked with two

A. tauschii haplotypes (Hap_AeT* and Hap_AeT), whereas the GWAS alleles associated with decreasing effects were in LD with Hap_HW contributed by the hexaploid wheat parents. The Hap_AeT* group of haplotypes was contributed by the *A. tauschii* parents having Hap2 and Hap7 at the *Ppd-D1* locus.

The Phenotypic Effects of Haplotype Block Variants

Average SNS and HD

Significant haplotype-trait associations were identified on chromosome arms 2DS and 4DL that influence SNS and HD. Chromosome 2DS had multiple introgressed haplotypes that are significantly associated with a variation in SNS and HD, with the most significant haplotypes located at 16,497,666–16,548,006 and 16,548,753–16,639,561 bp for SNS and HD, respectively. The haplotype variants with the increasing effect at these loci

TABLE 4 | ANOVA for the effect of *Ppd-D1* haplotype variants from hexaploid wheat and *A. tauschii* on the spikelet number per spike (SNS) in three experimental trials of the introgression population.

Haplotype	Estimate	Std. error	t-value	Pr(> t)
Hexaploid wheat (HW)	13.38	0.13	−3.41	0.000677***
<i>A. tauschii</i> (AeT)	13.83	0.12	115.84	<2e-16***
HW_Hap1	13.38	0.84	1.79	0.07317
AeT_Hap2	13.79	0.94	2.05	0.04092*
AeT_Hap3	13.74	0.91	2.06	0.03958*
AeT_Hap4	13.50	1.19	1.37	0.17002
AeT_Hap5	13.53	0.92	1.80	0.07190
AeT_Hap6	13.69	0.93	1.96	0.05069
AeT_Hap7	14.62	0.92	2.99	0.00288**
AeT_Hap8	13.77	1.19	1.60	0.11051

*** = 0.001, ** = 0.01, and * = 0.05.

Bold values indicate the haplotype that shows the highest significant increase in the spikelet number.

were from *A. tauschii* parents, whereas those with a reducing effect were from the hexaploid wheat lines (Figures 5C,D). The verification of GWAS results for an allelic effect at 2DS locus associated with SNS and HD supports the abovementioned observation (Supplementary File 8). We observed a positive Pearson's correlation coefficient between SNS and HD; and lines having haplotypes from either parent showed significant differences in the phenotype based on a *t*-test ($r = 0.23$, $p = 3.31 \times 10^{-07}$) at 95% confidence level. Haplotypes on 4DL had a smaller effect on SNS compared to HD. Among 35 and 66 ILs having SNS and HD trait values above the 90th percentile of trait distribution, respectively, 13 lines had the increasing alleles from *A. tauschii* at both 2DS and 4DL loci associated with SNS and HD traits (Figure 5E).

Initially, we did not detect a significant GWAS signal directly associated with SNPs within the *Ppd-D1* gene due to the lack of high-quality imputed SNPs in this region. Further analysis of parental haplotypes identified SNP variants linked with both the *Ppd-D1* haplotypes and haplotypes at 28 and 30 Mb region, showing a significant haplotype-trait association in two trials. These haplotypes were within ~3 Mb from the *Ppd-D1* locus and likely overlap with *Ppd-D1*. We performed the ANOVA to determine the effect of different haplotype variants identified in the parental accessions on the SNS in the introgression population using data from the three experimental trials (COI19, COI9, and AS20) (Table 4). The results show that both hexaploid and *A. tauschii* haplotypes have a significant effect on SNS ($p < 0.001$; Table 4, Figure 5B). Among the *A. tauschii* haplotypes, Hap7 had the highest impact on SNS ($p = 0.003$) followed by Hap2 ($p = 0.041$) and Hap3 ($p = 0.040$). In contrast, Hap5 with the His16Asn MS mutation had a negative effect on SNS and was not significantly different from Hap1 present in hexaploid wheat lines ($p = 0.072$). Consistent with previous studies, these results showed that the *Ppd-D1* gene located at ~34 Mb (33,961,438–33,951,651 bp interval in CS RefSeq v.1.0), which plays a role in flowering time regulation in wheat, also has a strong effect on the variation in the SNS (Beales et al., 2007; Guo et al., 2009).

TABLE 5 | Chromosome 2D haplotypes variants associated with the spikelet number and HD and how they influence other traits in the introgression population.

Trait	Haplotype	Loc1 Freq.	chr2D:16548753–16639561	SD
HD	Hap_HW	53	199 ^b	2.28*
HD	Hap_AeT	156	202^a	2.86[†]
HD	Hap_HW/AeT	142	200 ^c	2.29 [‡]
PH	Hap_HW	53	68.27 ^{ab}	6.79
PH	Hap_AeT	156	70.82 ^a	6.97
PH	Hap_HW/AeT	142	65.9 ^b	8.68
BM	Hap_HW	53	155.18 ^a	28.21
BM	Hap_AeT	156	156.26 ^a	31.01
BM	Hap_HW/AeT	142	157.52 ^a	35.82
SNS	Hap_HW	53	14.72 ^b	1.2
SNS	Hap_AeT	156	15.27^a	1.12
SNS	Hap_HW/AeT	142	14.05 ^c	1.03
GSW	Hap_HW	53	66.84 ^a	14.71
GSW	Hap_AeT	156	65.82 ^a	13.67
GSW	Hap_HW/AeT	142	66.47 ^a	13.57
HI	Hap_HW	53	42.85 ^a	4.35
HI	Hap_AeT	156	42.23 ^a	4.24
HI	Hap_HW/AeT	142	42.89 ^a	6.28

*Haplotype with a reducing effect, [†]Haplotype with an increasing effect, [‡]Haplotype with an intermediate effect, Loc1 = chr2D: 16,575,276–16,635,669 bp; Loc1 Freq. = Number of introgression lines having the haplotype; HD is a heading date, PH is an average plant height, BM is the aboveground total dry biomass, SNS is a spikelet number per spike, GSW is the grain sample weight from a biomass sample, and HI is the harvest index. Means with the same superscript letters are not statistically significant for each trait based on Tukey's honestly significant difference at 95% confidence level.

Bold values indicate the haplotype that shows the highest increase in both the SNS and HD.

Pleiotropic Effects of Haplotypes on Yield Component Traits

We also evaluated the effects of distinct haplotypes associated with HD on other traits. Haplotype Hap_AeT from chromosome 2D located at 16,548,753–16,639,561 bp is associated with a significant increase in the days to heading and SNS without any significant impact on BM, HI, and GSW (Table 5).

We compared the effects of two haplotypes associated with GW on 2DS (2D: 65,964,778–66,124,103 bp and 2D: 66,265,325–66,266,089 bp), where Hap_HW from hexaploid parents and Hap_AeT from *A. tauschii* ssp. *strangulata* increase GW, and Hap_AeT_{ta} from the *A. tauschii* ssp. *tauschii* parents reduces GW (Table 6). In Colby 2018 non-irrigated trial data, the haplotype at the 2D: 65,964,778–66,124,103 bp locus that was associated with an increase in GW and GL was also linked with an increase in GA and a decrease in GN. Meanwhile, both Hap_AeT_{st} and Hap_AeT_{ta} haplotypes at 2D: 65,964,778–66,124,103 bp were associated with an increase in GL, only Hap_AeT_{ta} was linked with a significant increase in GN (Table 6). The Hap_AeT_{ta} haplotype at the 2D: 66,265,325–66,266,089 bp haplotype block had similar effects on GN although the observed difference was not significant.

Similarly, haplotype block 6D: 463,775,852–463,809,722 bp associated with PH and SNS has three variants (Supplementary File 7). Haplotype variant Hap1_HW&AeT is associated with an increase in SNS, PH, BM, and GY. It showed

TABLE 6 | Chromosome 2D haplotypes variants associated with the grain width (GW) in Colby 2018 rainfed trial and their effects on other traits in the introgression population.

Trait	Haplotype	Haploblock chr2D:65964778–66124103			Haploblock chr2D:66265325–66266089		
		Freq.	Trait mean	Trait SD	Freq.	Trait mean	Trait SD
GN	Hap_AeT_st	5	475.50 ^b	27.86	20	504.15 ^a	55.28
GN	Hap_HW	140	499.04 ^b	44.3	304	502.97 ^a	41.6
GN	Hap_AeT_ta	202	510.94 ^a	43.38	27	540.37 ^a	47.94
TGW	Hap_AeT_st	5	32.74 ^a	2.15	20	31.54 ^a	2.74
TGW	Hap_HW	140	31.40 ^a	2.2	304	31.30 ^a	2
TGW	Hap_AeT_ta	202	31.01 ^a	1.98	27	29.54 ^a	1.76
GA	Hap_AeT_st	5	14.38 ^a	1.18	20	13.89 ^a	1.03
GA	Hap_HW	140	13.50 ^b	0.69	304	13.53 ^a	0.66
GA	Hap_AeT_ta	202	13.53 ^b	0.66	27	13.28 ^a	0.61
GW	Hap_AeT_st	5	3.14 ^a	0.05	20	3.09 ^{ab}	0.08Δ
GW	Hap_HW	140	3.13 ^a	0.09	304	3.09 ^a	0.09Δ
GW	Hap_AeT_ta	202	3.06 ^b	0.08	27	3.00 ^b	0.07#
GL	Hap_AeT_st	5	6.25 ^a	0.51	20	6.11 ^a	0.4
GL	Hap_HW	140	5.85 ^b	0.22	304	5.93 ^b	0.24
GL	Hap_AeT_ta	202	6.00 ^a	0.24	27	6.01 ^a	0.22
GY	Hap_AeT_st	5	59.19 ^a	14.46	20	58.76 ^a	14.77
GY	Hap_HW	140	60.51 ^a	12.04	304	61.26 ^a	12.03
GY	Hap_AeT_ta	202	61.28 ^a	12.54	27	59.55 ^a	13.41

Δhaplotype that increases GW, # haplotype that reduces GW. Means with the same superscript letters are not statistically significant for each trait based on Tukey's honestly significant difference at 95% confidence level.

no association with HD and grain traits, except GN where an intermediate effect was observed. This haplotype variant is present in two *A. tauschii* lines (TA2389 and TA2398) and the hexaploid parents excluding KanMark and KS061862M-17. Haplotype variant Hap0_HW&AeT, which is found in KanMark and KS061862M-17 and two ssp. *strangulata* accessions (TA1642 and TA2378), has an intermediate effect on GY, reducing it by two bushels compared to Hap1_HW&AeT. The third haplotype variant (Hap_AeT) is present only in the *A. tauschii* lines.

The haplotype Hap_AeT contributed by *A. tauschii* at haplotype block 7D: 14,185,651–14,596,748 bp, was associated with a significant increase in SNS compared to haplotypes present in winter wheat (Supplementary Table 2). This increase was associated with a significant decrease in GL and had no significant effect on GY. At 7D: 14,722,457–14,817,138 bp, the Hap_AeT haplotype contributed by *A. tauschii* was also linked with a significant increase in SNS compared to Hap1_HW&AeT detected in both hexaploid wheat and *A. tauschii* parents. However, an increase in SNS for this haplotype was connected with a decrease in both GL and GY. At this haplotype block (7D: 14,722,457–14,817,138 bp), the most significant increase in GY was observed for lines carrying haplotype Hap0_HW&AeT, which was associated with a moderate increase in both SNS and GL (Supplementary Table 2).

DISCUSSION

Here, we performed the sequence-based analysis of haplotypes in the wild-relative introgression population developed by crossing a diverse panel of *A. tauschii* accessions with winter wheat cultivars. Our results demonstrate that the whole-genome sequencing of wild and cultivated wheat founder

lines in combination with the complexity-reduced sequencing of a derived introgression population provides an effective framework for SNP imputation. Because most breeding populations are based on a limited number of founders, often including 10–30 lines, their whole-genome sequencing is feasible in crops even with large genomes such as wheat, and provides a comprehensive description of allelic diversity present in a breeding population. The latter makes sequenced founders an ideal reference panel for imputing genotypes in a breeding population genotyped using low-coverage or complexity-reduced sequencing. This was recently demonstrated by imputing genotypes in the wheat MAGIC population genotyped by low-coverage sequencing (Scott et al., 2021). The composition of our introgression population, including multiple biparental cross families (Nyine et al., 2020), also shifts the population allele frequency toward more common variants, which could be imputed with a higher accuracy than rare variants (Huang et al., 2015). In addition, the high levels of LD in the introgression population should increase the length of haplotype blocks and facilitate the detection of matching haplotypes in the reference panel of founders using even sparse genotyping data generated by low-coverage or complexity-reducing sequencing. Consistent with these assumptions, an imputation algorithm implemented in Beagle (Browning and Browning, 2013) allowed us to impute nearly 5.2 million SNPs in the introgression population with high genotype call probabilities above 0.7 using SNPs generated by complexity-reduced sequencing of this population and nearly 20 million variants identified in 27 founders. This high-density SNP marker data permitted a detailed characterization of the introgressed haplotypes (Pook et al., 2019) and assessing their effects on productivity traits.

Our results demonstrate that wild-relative introgressions into the D genome of wheat, the least diverse amongst the three subgenomes (Chao et al., 2010; Jordan et al., 2015; Singh et al., 2019), is associated with the increased levels of variation in yield and yield component traits. The analysis of data from several years and locations under irrigated and non-irrigated conditions revealed many superior ILs that produce more grains or show higher yield stability than the control cultivars. The yield increase in top-performing ILs was driven by a combination of yield component traits, and in many cases, it was associated with increased grain size, grain weight, and BM or improved harvest index. These results suggest that wild-relative introgression has the potential to positively affect source-sink balance, which was suggested to be one of the important factors contributing to yield potential (Reynolds et al., 2017). Many of these high yielding lines (~23%) were also among the top lines showing the highest levels of yield stability, indicating that the introgression from *A. tauschii* likely improves the adaptive potential of hard red winter wheat in different environmental conditions. Consistent with this conclusion, the highest impact of introgression on yield was found in a non-irrigated trial, indicating that alleles from *A. tauschii* likely improve the adaption of hexaploid wheat to water-limiting conditions. The *A. tauschii* accessions used to generate the introgression population represent both L1 and L2 lineages (Wang et al., 2013) and originate from a broad range of geographical locations with variable climatic conditions, likely capturing adaptive haplotypes from the regions prone to drought stress.

Heading date is one of the key agronomic traits linked with wheat adaptation to different geographical locations and improvement in yield (Jung and Müller, 2009). In our population, a positive correlation was observed between HD and the SNS, with some lines showing up to 2-week delay in HD. Several haplotype blocks on chromosome arms 2DS and 4DL were significantly associated with a variation in spikelet number and HD. The haplotypes with increasing effects at both loci were derived from *A. tauschii*, indicating their potential for modulating both traits in bread wheat. Chromosome 2DS is known to harbor the *Ppd-D1* and *Rht8* genes that control flowering time and PH, respectively, and also could affect the spikelet number (Shaw et al., 2013; Muqaddasi et al., 2019). The overlapping haplotype blocks associated with the spikelet number and HD were identified on 2DS, confirming that the two traits co-segregating in the population have a common genetic basis. We demonstrated that these 2DS haplotypes are associated with the different allelic variants of the *Ppd-D1* gene from *A. tauschii*. Consistent with the earlier studies, these results demonstrated that the different alleles of the *Ppd-D1* gene have distinct effects on HD and SNS (Beales et al., 2007; Guo et al., 2009). These effects were correlated with the relative expression levels of each *Ppd-D1* allele (Guo et al., 2009), suggesting that functional mutations within the *Ppd-D1* coding region and the modifier mutations in the regulatory region of the gene likely to account for a variation in these traits in the *A. tauschii* winter wheat introgression population. The developmental plasticity modulated by *Ppd-D1* is mediated by the changes in the expression of flowering time genes (Gol et al., 2021). It was

shown that the *Ppd-H1* from wild barley is capable of integrating environmental signals to control HD and minimize the negative impact of transient drought stress on spikelet number (Gol et al., 2021). Consistent with this observation, in the current study, ILs that have a high proportion of *A. tauschii* segments produced more grain under drought stress in the Colby 2018 trial, raising the possibility that the *A. tauschii* alleles of *Ppd-D1* also have the potential to protect wheat from the physiological effects of stress that lead to low yield.

Our study reveals that some haplotypes associated with the productivity trait variation in the introgression population also exhibit significant pleiotropic effects. Meanwhile, the direction of effects on various traits was largely consistent with the previously reported trade-offs among component traits (Griffiths et al., 2015; Reynolds et al., 2017; Quintero et al., 2018), the combined effects of some introgressed haplotypes were associated with the positive trends in yield. For example, a haplotype contributed by *A. tauschii* ssp. *tauschii* at the chromosome 2D haplotype block at 65,964,778–66,124,103 bp was associated with an increase in GL, size, and number with a moderate positive effect on GY. At the haplotype block on chromosome 7D located between 14,722,457–14,817,138 bp, the Hap0_HW&AeT haplotype shared between hexaploid wheat and *A. tauschii* parents and associated with a moderate increase in both SNS and GL was also associated with the most significant increase in GY. Analyses of the pleiotropic effects of the introgressed haplotypes suggest that these haplotypes on chromosomes 2D and 7D could be utilized in breeding programs to improve yield component traits without negative effects on other productivity traits.

CONCLUSIONS

The imputation of markers from whole-genome-sequenced reference panels into skim-sequenced inference populations is increasingly becoming a common practice in plant breeding program due to its cost-effectiveness (Happ et al., 2019; Jensen et al., 2020). Our study demonstrates the utility of this strategy for detecting introgression in the wheat genome and contributes to developing genomic resources for deploying wild-relative diversity in wheat breeding programs. We show that the haplotype-based analysis of trait variation in this population has the potential to improve our knowledge on the genetic effects of the introgressed diversity on productivity traits and identify novel haplotypes for improving yield potential in wheat.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

AF and MC generated the *A. tauschii* introgression population. MN, EAd, MC, RA, BB, WW, DD, ZY, YG, FH, KJ, and AF

phenotyped the introgression population. MN, BB, and AA generated the genotyping data. AA performed next-generation sequencing of parental lines and introgression population. MN and FH performed the bioinformatical analyses of data. MN performed the statistical analysis. EAK conceived the idea and interpreted results. MN and EAK wrote the manuscript. All authors contributed to the manuscript revision, read, and approved the submitted version.

FUNDING

This research was supported by the Agriculture and Food Research Initiative Competitive Grants 2017-67007-25939 (WheatCAP) and 2020-68013-30905 from the USDA National

Institute of Food and Agriculture, and by the International Wheat Yield Partnership (IWYP).

ACKNOWLEDGMENTS

We would like to thank Mary Guttieri for providing access to the USDA ARS seed processing facility.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.716955/full#supplementary-material>

REFERENCES

- Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E. D., and Schroeder, J. I. (2019). Genetic strategies for improving crop yields. *Nature*. 575, 109–118. doi: 10.1038/s41586-019-1679-0
- Beales, J., Turner, A., Griffiths, S., Snape, J. W., and Laurie, D. A. (2007). A pseudo-response regulator is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 115, 721–733. doi: 10.1007/s00122-007-0603-4
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Borojevic, K., and Borojevic, K. (2005). The transfer and history of “reduced height genes” (Rht) in Wheat from Japan to Europe. *J. Hered.* 96, 455–459. doi: 10.1093/jhered/esi060
- Browning, B. L., and Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* 194, 459–471. doi: 10.1534/genetics.113.150029
- Chao, S., Dubcovsky, J., Dvorak, J., Luo, M.-C., Baenziger, S. P., Matnyazov, R., et al. (2010). Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC Genom.* 11:727. doi: 10.1186/1471-2164-11-727
- Chen, L., Du, Y., Lu, Q., Chen, H., Meng, R., Cui, C., et al. (2018). The photoperiod-insensitive allele *Ppd-D1a* promotes earlier flowering in Rht12 dwarf plants of bread wheat. *Front. Plant Sci.* 9:1312. doi: 10.3389/fpls.2018.01312
- Dale, Z., Jie, H., Luyu, H., Cancan, Z., Yun, Z., Yaru, S., et al. (2017). An advanced backcross population through synthetic octaploid wheat as a “bridge”: development and QTL detection for seed dormancy. *Front. Plant Sci.* 8:2123. doi: 10.3389/fpls.2017.02123
- Davies, R. W., Flint, J., Myers, S., and Mott, R. (2016). Rapid genotype imputation from sequence without reference panels. *Nat. Genet.* 48, 965–969. doi: 10.1038/ng.3594
- Del Moral, L. G., Rharrabti, Y., Villegas, D., and Royo, C. (2003). Evaluation of grain yield and its components in durum wheat under Mediterranean conditions: an ontogenic approach. *Agron. J.* 95, 266–274. doi: 10.2134/agronj2003.0266
- Du, Y., Cui, B., Wang, Z., Sun, J., and Niu, W. (2020). Effects of manure fertilizer on crop yield and soil properties in China: a meta-analysis. *Catena* 193:104617. doi: 10.1016/j.catena.2020.104617
- Dvorák, J., Edge, M., and Ross, K. (1988). On the evolution of the adaptation of *Lophopyrum elongatum* to growth in saline environments. *PNAS* 85, 3805–3809. doi: 10.1073/pnas.85.11.3805
- Gol, L., Haraldsson, E. B., and von Korff, M. (2021). *Ppd-H1* integrates drought stress signals to control spike development and flowering time in barley. *J. Exp. Bot.* 72, 122–136. doi: 10.1093/jxb/eraa261
- Golan, G., Ayalon, I., Perry, A., Zimran, G., Ade-Ajayi, T., Mosquna, A., et al. (2019). GNI - A1 mediates trade-off between grain number and grain weight in tetraploid wheat. *Theor. Appl. Genet.* 132, 2353–2365. doi: 10.1007/s00122-019-03358-5
- Griffiths, S., Wingen, L., Pietragalla, J., Garcia, G., Hasan, A., Miralles, D., et al. (2015). Genetic dissection of grain size and grain number trade-offs in CIMMYT wheat germplasm. *PLoS ONE* 10:e0118847. doi: 10.1371/journal.pone.0118847
- Guo, Z., Song, Y., Zhou, R., Ren, Z., and Jia, J. (2009). Discovery, evaluation and distribution of haplotypes of the wheat *Ppd-D1* gene. *New Phytol.* 185, 841–851. doi: 10.1111/j.1469-8137.2009.03099.x
- Hao, M., Zhang, L., Ning, S., Huang, L., Yuan, Z., Wu, B., et al. (2020). The resurgence of introgression breeding, as exemplified in wheat improvement. *Front. Plant Sci.* 11:252. doi: 10.3389/fpls.2020.00252
- Happ, M. M., Wang, H., Graef, G. L., and Hyten, D. L. (2019). Generating High density, low cost genotype data in soybean [*Glycine max* (L.) Merr.]. *G3 Genes Genomes Genet.* 9, 2153–2160. doi: 10.1534/g3.119.400093
- Hill, W. G., and Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genet. Res.* 8, 269–294. doi: 10.1017/S0016672300010156
- Huang, J., Howie, B., McCarthy, S., Memari, Y., Walter, K., Min, J. L., et al. (2015). Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat. Commun.* 6:8111. doi: 10.1038/ncomms9111
- International Wheat Genome Sequencing Consortium (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361, 1–16. doi: 10.1126/science.aar7191
- Jensen, S. E., Charles, J. R., Muleta, K., Bradbury, P. J., Casstevens, T., Deshpande, S. P., et al. (2020). A sorghum practical haplotype graph facilitates genome-wide imputation and cost-effective genomic prediction. *Plant Genome* 13:e20009. doi: 10.1002/tpg2.20009
- Jones, B. H., Blake, N. K., Heo, H.-Y., Kalous, J. R., Martin, J. M., Torrión, J. A., et al. (2020). Improving hexaploid spring wheat by introgression of alleles for yield component traits from durum wheat. *Crop Sci.* 60, 759–771. doi: 10.1002/csc2.20011
- Jordan, K. W., Wang, S., Lun, Y., Gardiner, L.-J., MacLachlan, R., Hucl, P., et al. (2015). A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biol.* 16:48. doi: 10.1186/s13059-015-0606-4
- Jung, C., and Müller, A. E. (2009). Flowering time control and applications in plant breeding. *Trends Plant Sci.* 14, 563–573. doi: 10.1016/j.tplants.2009.07.005
- Kuzay, S., Xu, Y., Zhang, J., Katz, A., Pearce, S., Su, Z., et al. (2019). Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. *Theor. Appl. Genet.* 132, 2689–2705. doi: 10.1007/s00122-019-03382-5
- Li, Y., Willer, C. J., Ding, J., Scheet, P., and Abecasis, G. R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* 34, 816–834. doi: 10.1002/gepi.20533
- Lorenz, A. J., Hamblin, M. T., and Jannink, J.-L. (2010). Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in Barley. *PLoS ONE* 5:e14079. doi: 10.1371/journal.pone.0014079

- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Multize, D. K. (1990). AGROBASE/4: a microcomputer database management and analysis system for plant breeding and agronomy. *Agron. J.* 82, 1016–1021. doi: 10.2134/agronj1990.00021962008200050035x
- Muqaddasi, Q. H., Brassac, J., Koppolu, R., Plieske, J., Ganal, M. W., and Röder, M. S. (2019). TaAPO-A1, an ortholog of rice aberrant panicle organization 1, is associated with total spikelet number per spike in elite European hexaploid winter wheat (*Triticum aestivum* L.) varieties. *Sci. Rep.* 9:13853. doi: 10.1038/s41598-019-50331-9
- Nyine, M., Adhikari, E., Clinesmith, M., Jordan, K. W., Fritz, A. K., and Akhunov, E. (2020). Genomic patterns of introgression in interspecific populations created by crossing wheat with its wild relative. *G3 Genes Genomes Genet.* 10, 3651–3661. doi: 10.1534/g3.120.401479
- Nyine, M., Wang, S., Kiani, K., Jordan, K., Liu, S., Byrne, P., et al. (2019). Genotype imputation in winter wheat using first-generation haplotype map SNPs improves genome-wide association mapping and genomic prediction of traits. *G3 Genes Genomes Genet.* 9, 125–133. doi: 10.1534/g3.118.200664
- Pasquariello, M., Berry, S., Burt, C., Uauy, C., and Nicholson, P. (2020). Yield reduction historically associated with the *Aegilops ventricosa* 7D V introgression is genetically and physically distinct from the eyespot resistance gene Pch1. *Theor. Appl. Genet.* 133, 707–717. doi: 10.1007/s00122-019-03502-1
- Peleg, Z., Fahima, T., Abbo, S., Krugman, T., Nevo, E., Yakir, D., et al. (2005). Genetic diversity for drought resistance in wild emmer wheat and its ecogeographical associations. *Plant Cell Environ.* 28, 176–191. doi: 10.1111/j.1365-3040.2005.01259.x
- Pennacchi, J. P., Carmo-Silva, E., Andralojc, P. J., Lawson, T., Allen, A. M., Raines, C. A., et al. (2019). Stability of wheat grain yields over three field seasons in the UK. *Food Energy Secur.* 8:e00147. doi: 10.1002/fes3.147
- Periyannan, S., Moore, J., Ayliffe, M., Bansal, U., Wang, X., Huang, L., et al. (2013). The gene Sr33, an ortholog of barley mla genes, encodes resistance to wheat stem rust race Ug99. *Science* 341, 786–788. doi: 10.1126/science.1239028
- Placido, D. F., Campbell, M. T., Folsom, J. J., Cui, X., Kruger, G. R., Baenziger, P. S., et al. (2013). Introgression of novel traits from a wild wheat relative improves drought adaptation in wheat. *Plant Physiol.* 161, 1806–1819. doi: 10.1104/pp.113.214262
- Pook, T., Schlather, M., Campos, G., de los, Mayer, M., Schoen, C. C., and Simianer, H. (2019). HaploBlocker: creation of subgroup-specific haplotype blocks and libraries. *Genetics* 212, 1045–1061. doi: 10.1534/genetics.119.302283
- Pour-Aboughadareh, A., Yousefian, M., Moradkhani, H., Pocai, P., and Siddique, K. H. M. (2019). STABILITYSOFT: a new online program to calculate parametric and non-parametric stability statistics for crop traits. *Appl. Plant Sci.* 7:e01211. doi: 10.1002/aps3.1211
- Quintero, A., Molero, G., Reynolds, M. P., and Calderini, D. F. (2018). Trade-off between grain weight and grain number in wheat depends on GxE interaction: a case study of an elite CIMMYT panel (CIMCOG). *Eur. J. Agron.* 92, 17–29. doi: 10.1016/j.eja.2017.09.007
- Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., et al. (2020). Breeder friendly phenotyping. *Plant Sci.* 295:110396. doi: 10.1016/j.plantsci.2019.110396
- Reynolds, M. P., Calderini, D. F., Condon, A. G., and Rajaram, S. (2001). Physiological basis of yield gains in wheat associated with the LR19 translocation from *Agropyron elongatum*. *Euphytica* 119, 139–144. doi: 10.1023/A:1017521800795
- Reynolds, M. P., Pask, A. J. D., Hoppitt, W. J. E., Sonder, K., Sukumaran, S., Molero, G., et al. (2017). Strategic crossing of biomass and harvest index—source and sink—achieves genetic gains in wheat. *Euphytica* 213, 1–23. doi: 10.1007/s10681-017-2040-z
- Saintenac, C., Zhang, W., Salcedo, A., Rouse, M. N., Trick, H. N., Akhunov, E., et al. (2013). Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. *Science* 341, 783–786. doi: 10.1126/science.1239022
- Scott, M. F., Fradgley, N., Bentley, A. R., Brabbs, T., Corke, F., Gardner, K. A., et al. (2021). Limited haplotype diversity underlies polygenic trait architecture across 70 years of wheat breeding. *Genome Biol.* 22:137. doi: 10.1186/s13059-021-02354-7
- Shaw, L. M., Turner, A. S., Herry, L., Griffiths, S., and Laurie, D. A. (2013). Mutant alleles of photoperiod-1 in wheat (*Triticum aestivum* L.) that confer a late flowering phenotype in long days. *PLoS ONE* 8:e79459. doi: 10.1371/journal.pone.0079459
- Singh, N., Wu, S., Tiwari, V., Sehgal, S., Raupp, J., Wilson, D., et al. (2019). Genomic analysis confirms population structure and identifies inter-lineage hybrids in *Aegilops tauschii*. *Front. Plant Sci.* 10:9. doi: 10.3389/fpls.2019.00009
- Swarts, K., Li, H., Romero Navarro, J. A., An, D., Romay, M. C., Hearne, S., et al. (2014). Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *Plant Genome* 7, 1–12. doi: 10.3835/plantgenome2014.05.0023
- Wang, J., Luo, M. C., Chen, Z., You, F. M., Wei, Y., Zheng, Y., et al. (2013). *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. *New Phytol.* 198, 925–937. doi: 10.1111/nph.12164
- Wang, R. R.-C., and Zhang, X.-Y. (1996). Characterization of the translocated chromosome using fluorescence in situ hybridization and random amplified polymorphic DNA on two *Triticum aestivum*-thinopyrum intermedium translocation lines resistant to wheat streak mosaic or barley yellow dwarf virus. *Chromosome Res.* 4, 583–587. doi: 10.1007/BF02261721
- Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796. doi: 10.1111/pbi.12183
- Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., and Song, B.-H. (2017). Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement. *Evol. Appl.* 10, 5–24. doi: 10.1111/eva.12434
- Zhang, J., Gizaw, S. A., Bossolini, E., Hegarty, J., Howell, T., Carter, A. H., et al. (2018). Identification and validation of QTL for grain yield and plant water status under contrasting water treatments in fall-sown spring wheats. *Theor. Appl. Genet.* 131, 1741–1759. doi: 10.1007/s00122-018-3111-9
- Zhang, K., Calabrese, P., Nordborg, M., and Sun, F. (2002). Haplotype block structure and its applications to association studies: power and study designs. *Am. J. Hum. Genet.* 71, 1386–1394. doi: 10.1086/344780

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Nyine, Adhikari, Clinesmith, Aiken, Betzen, Wang, Davidson, Yu, Guo, He, Akhunova, Jordan, Fritz and Akhunov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Multi-Trait Multi-Environment Genomic Prediction of Agronomic Traits in Advanced Breeding Lines of Winter Wheat

Harsimardeep S. Gill¹, Jyotirmoy Halder¹, Jinfeng Zhang¹, Navreet K. Brar¹, Teerath S. Rai¹, Cody Hall¹, Amy Bernardo², Paul St Amand³, Guihua Bai³, Eric Olson⁴, Shaikat Ali¹, Brent Turnipseed¹ and Sunish K. Sehgal^{1*}

¹ Department of Agronomy, Horticulture & Plant Science, South Dakota State University, Brookings, SD, United States,

² Department of Plant Pathology, Kansas State University, Manhattan, KS, United States, ³ United States Department of Agriculture - Agricultural Research Services, Hard Winter Wheat Genetic Research Unit, Manhattan, KS, United States,

⁴ Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI, United States

OPEN ACCESS

Edited by:

Hakan Ozkan,
Çukurova University, Turkey

Reviewed by:

Li Li,
University of New England, Australia
Ilaria Marcotuli,
University of Bari Aldo Moro, Italy

*Correspondence:

Sunish K. Sehgal
sunish.sehgal@sdstate.edu

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 14 May 2021

Accepted: 12 July 2021

Published: 18 August 2021

Citation:

Gill HS, Halder J, Zhang J, Brar NK, Rai TS, Hall C, Bernardo A, Amand PS, Bai G, Olson E, Ali S, Turnipseed B and Sehgal SK (2021) Multi-Trait Multi-Environment Genomic Prediction of Agronomic Traits in Advanced Breeding Lines of Winter Wheat. *Front. Plant Sci.* 12:709545. doi: 10.3389/fpls.2021.709545

Genomic prediction is a promising approach for accelerating the genetic gain of complex traits in wheat breeding. However, increasing the prediction accuracy (PA) of genomic prediction (GP) models remains a challenge in the successful implementation of this approach. Multivariate models have shown promise when evaluated using diverse panels of unrelated accessions; however, limited information is available on their performance in advanced breeding trials. Here, we used multivariate GP models to predict multiple agronomic traits using 314 advanced and elite breeding lines of winter wheat evaluated in 10 site-year environments. We evaluated a multi-trait (MT) model with two cross-validation schemes representing different breeding scenarios (CV1, prediction of completely unphenotyped lines; and CV2, prediction of partially phenotyped lines for correlated traits). Moreover, extensive data from multi-environment trials (METs) were used to cross-validate a Bayesian multi-trait multi-environment (MTME) model that integrates the analysis of multiple-traits, such as $G \times E$ interaction. The MT-CV2 model outperformed all the other models for predicting grain yield with significant improvement in PA over the single-trait (ST-CV1) model. The MTME model performed better for all traits, with average improvement over the ST-CV1 reaching up to 19, 71, 17, 48, and 51% for grain yield, grain protein content, test weight, plant height, and days to heading, respectively. Overall, the empirical analyses elucidate the potential of both the MT-CV2 and MTME models when advanced breeding lines are used as a training population to predict related preliminary breeding lines. Further, we evaluated the practical application of the MTME model in the breeding program to reduce phenotyping cost using a sparse testing design. This showed that complementing METs with GP can substantially enhance resource efficiency. Our results demonstrate that multivariate GS models have a great potential in implementing GS in breeding programs.

Keywords: BMTME, GBS, genomic prediction, genomic selection, $G \times E$, multi-trait multi-environment genomic prediction, wheat breeding

INTRODUCTION

Global wheat production needs to be increased by 60% to meet the demand of a projected population of 9 billion by 2050 (Tester and Langridge, 2010; Fischer et al., 2014). In the past few decades, wheat breeding successfully achieved a significant increase in grain yield owing to significantly improved genetic resources, implementation of modern agronomic practices, accurate experimental designs, and other improved technology packages (Tadesse et al., 2019), which translates into an annual increase of 1% in terms of genetic gain in grain yield. However, this increase is still far from the expected yearly growth of 1.7% to meet the future wheat demand (Oury et al., 2012; Tadesse et al., 2019). Thus, new and innovative breeding technologies are essential to achieve a 2-fold increase in annual yield to avoid potential food crises in the coming decades.

Traditional wheat breeding involves creating novel genetic variation with different methods, followed by extensive selection and advancement of generations. The selection of progeny with desirable agronomic and end-use quality traits is a resource-intensive process and could take up to 10–15 years to develop a new cultivar (Haile et al., 2020). Further, in traits with complex genetic architecture such as grain yield, genotype-by-environment interactions play a paramount role and impose additional challenges in selection. In recent years, the deployment of molecular markers for marker-assisted selection (MAS) has been used to increase the selection accuracy and accelerate genetic gain (Randhawa et al., 2013). Although MAS has shown a good potential in wheat breeding for the deployment of qualitative trait loci (QTLs) with large effects, its application has been limited to improve complex traits governed by many QTLs with small effects (Heffner et al., 2009).

Genomic selection (GS) is a recent approach that utilizes genome-wide marker data to select individuals superior for complex traits in the early breeding cycle to increase the genetic gain per unit of time (Meuwissen et al., 2001; Heffner et al., 2009). Unlike MAS, GS does not require prior identification of QTLs for traits of interest; instead, it employs all available markers across the genome to predict breeding values of individuals (Bassi et al., 2015). Briefly, GS requires a training population (TP), which is genotyped with genome-wide markers and for a given trait(s) of interest. GS involves the calibration of a prediction model using the TP to estimate marker effects and evaluate the predictive ability of the model through cross-validation. Finally, the developed model is used to calculate genome-estimated breeding values (GEBVs) and rank the lines from a breeding or testing population (BP) that consists of lines with only genotypic information. Thus, the early selection or culling of individuals based on GEBVs permits greater genetic gain per breeding cycle, facilitating an increase in the efficacy of breeding programs and resulting in reduced varietal development costs. Several studies have reported the successful implementation of GS in different crops resulting in an accelerated rate of genetic gain compared with traditional breeding (Bassi et al., 2015; Battenfield et al., 2016; Bhat et al., 2016). Moreover, GS has shown to be particularly useful in traits where phenotyping is cumbersome, such as quality traits and

complex resistance to diseases (Battenfield et al., 2016; Dong et al., 2018).

The widespread availability of genome-wide markers attributed to low-cost genotyping technologies has facilitated the adaptability of GS in wheat breeding programs (Poland et al., 2012b; Bhat et al., 2016). Thus, in recent years, there has been a growing interest to complement phenotyping selection and genomic selection in wheat breeding. GS has been evaluated for many complex traits in wheat, including but not limited to grain yield and yield-related traits (Rutkoski et al., 2016; Ward et al., 2019; Guo et al., 2020; Haile et al., 2020; Juliana et al., 2020), wheat resistance to rusts (Rutkoski et al., 2014; Juliana et al., 2017), Fusarium head blight (Rutkoski et al., 2012; Arruda et al., 2015; Dong et al., 2018), and end-use quality traits (Battenfield et al., 2016; Lado et al., 2018; Ibba et al., 2020). Despite the successful evaluations of GS in wheat breeding programs, there is a continuous scope to improve the prediction accuracy/ability of GS models for quantitative traits to achieve higher genetic gains that will lead to the routine implementation of GS in various wheat breeding schemes.

The predictive ability of the genomic selection model refers to the correlation between estimated genome-estimated breeding values and the actual phenotypic values of individuals in a validation set and is generally calculated through a cross-validation approach. Along with TP size, extent of linkage disequilibrium (LD), and heritability of traits, predictive ability also depends on the choice and optimization of statistical models (de los Campos et al., 2013; Rutkoski et al., 2016; Guo et al., 2020). In most studies, penalized genomic prediction models, such as ridge-regression best linear unbiased prediction (rrBLUP) and genomic best linear unbiased prediction (GBLUP), have been standard GS approaches (VanRaden et al., 2009; Endelman, 2011). In addition, several Bayesian methods with different prior distributions and relying on Markov-Chain Monte Carlo (MCMC) for estimation of parameters have proven useful for genomic prediction (Habier et al., 2011; Wang et al., 2018). However, most of these models implement a univariate linear mixed model and are helpful in predicting only one variable at a time.

In recent years, multi-trait genomic prediction models have been suggested to improve the PA for a primary trait when secondary traits correlated to the primary trait are available (Jia and Jannink, 2012). The use of genetically correlated traits is of particular importance when the primary trait is difficult or expensive to phenotype and has low heritability. Several empirical studies have successfully evaluated multi-trait (MT) approaches for different agronomic traits in wheat breeding (Rutkoski et al., 2012; Hayes et al., 2017; Lado et al., 2018). An improvement of 70% in PA for grain yield was observed by including canopy temperature (CT) and normalized difference vegetation index as secondary traits using the MT approach (Rutkoski et al., 2016; Sun et al., 2017). Similarly, Hayes et al. (2017) and Lado et al. (2018) observed an increase in PA using multivariate approaches (MT) over single trait (ST) models in end-use quality traits.

For complex traits, genotype-by-environment ($G \times E$) interactions necessitate the evaluation of breeding lines for

multiple traits over multiple environments. Thus, the extension of MT approaches to account for a $G \times E$ interaction could improve the model for genomic prediction accuracy in breeding programs. Montesinos-López et al. (2016) proposed a Bayesian multi-trait and multi-environment (BMTME) model that integrates the analysis of multi-traits recorded over multi-environments in a unified approach. Recently, an improved BMTME model has been introduced that estimates the variance-covariance structure among traits, genotypes, and environments to predict multiple traits evaluated in various environments (Montesinos-López et al., 2019). Some studies using simulated and empirical data found that the BMTME model outperforms ST models in agronomic and end-use quality traits in wheat (Montesinos-López et al., 2016; Guo et al., 2020; Ibbá et al., 2020). The better performance of multivariate GS approaches stimulates us to evaluate these models in an actual breeding pipeline, where several traits are evaluated over diverse environments.

Although different GS approaches have been tested for predicting complex traits in wheat breeding programs, only few studies have reported the application of GS in actual yield trials where lines are evaluated over several environments (Belamkar et al., 2018). GS has a great potential in the early selection or culling in preliminary trials using information from advanced trials and accelerates genetic improvement. Furthermore, GS can complement phenotypic selection in practical scenarios such as loss of complete/partial trials due to weather extremes. In this study, we focused on the use of advanced breeding lines evaluated over multiple environments as training sets to predict untested genotypes using univariate and multivariate GS approaches. The specific objectives of this study were to (1) estimate the PA of various agronomic traits in advanced breeding lines using univariate and multivariate GP models and different cross-validation schemes, (2) assess the reliability of multivariate GP models in predicting complex traits over different years and locations, and (3) investigate the application of multi-trait multi-environment GP models in sparse testing of breeding lines.

MATERIALS AND METHODS

Plant Materials

The experiment was conducted for two growing seasons (2018–19 and 2019–20) using a total of 314 winter wheat genotypes. The genotypes included breeding lines from 2018 to 2019 and 2019 to 2020 wheat advanced yield trials (AYTs) and elite yield trials (EYTs) from the South Dakota State University (SDSU) winter wheat breeding program and well-adapted check cultivars. Most of the genotypes were either $F_{4:7}$ or $F_{4:8}$ filial generation. Of the 314 genotypes, 157 were evaluated in the growing season of 2019 and another 157 in that of 2020. Forty-four genotypes were shared between the two sets of wheat materials, leaving 270 unique genotypes in the study. We removed seven genotypes from genomic prediction analyses because of low-quality genotypic data. Thus, 151 and 156 genotypes were used for further analyses in the 2018–19 and 2019–20 growing seasons, respectively.

Experimental Design and Trait Measurement

The experimental plots were planted under a no-till system at five locations in South Dakota (**Supplementary Table 1**) in both seasons. The experimental unit at each of the five locations consisted of 1.5-m wide and 4-m long plots with seven rows spaced 20 cm apart. The seeding rate for plots was 300 seeds m^{-2} at all the locations. Recommended agronomic practices were followed for proper growth and yield.

Five agronomic traits were measured in this study, namely, grain yield (YLD) (bushels $acre^{-1}$), grain protein content (PROT) (%), test weight (TW) ($kg\ hL^{-1}$), plant height (HT) (cm), and days to heading (HDs) (Julian days). YLD was determined after harvesting the plots upon maturity using a plot combine (Zurn, Westernhausen Germany). PROT, TW, and moisture content were measured using Infratec™ 1241 Grain Analyzer (FOSS North America, Eden Prairie, MN, United States). YLD from plot and PROT were adjusted to 13% moisture content equivalence. HT was recorded as the distance from the soil surface to the tip of the fully emerged spike, excluding any awns if present. HDs were recorded as the Julian days required for 50% of heads to emerge from the boot in each plot.

Phenotypic Data Analysis

The phenotypic data for all the five agronomic traits were analyzed using best linear unbiased estimates (BLUEs) for individual environments. The model used for estimation of the genotypic BLUEs for individual environments was as follows:

$$y_{ij} = \mu + R_i + G_j + e_{ij} \quad (1)$$

where y_{ij} is the trait of interest, μ is the overall mean, R_i is the effect of the i^{th} replicate, G_j is the effect of the j^{th} genotype, and e_{ij} is the residual error effect associated with the i^{th} replication and j^{th} genotype. The replicates correspond to the complete blocks.

For the across environment estimation of best linear unbiased estimates (BLUEs) and best linear unbiased predictions (BLUPs), the statistical model was modified, as shown below:

$$y_{ijk} = \mu + E_i + R_{j(i)} + G_k + GE_{ik} + e_{ijk} \quad (2)$$

where y_{ijk} is the trait of interest, μ is the overall mean, E_i is the effect of the i^{th} environment, $R_{j(i)}$ is the effect of the j^{th} replicate nested within the i^{th} environment, G_k is the effect of the k^{th} genotype, GE_{ik} is the effect of the genotype $\times$ environment ($G \times E$) interaction, and e_{ijk} is the residual error effect associated with the i^{th} replication and j^{th} genotype. The environment corresponds to the individual locations and replicates correspond to the complete blocks. The genotype was assumed as a fixed effect, whereas the environment and block nested within the environment were assumed as random effects.

The broad-sense heritability (H^2) of a trait of interest in an independent environment was assessed as follows:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2 / nRep} \quad (3)$$

where σ_g^2 and σ_e^2 are the genotype and error variance components, respectively. The BLUEs and variance components were estimated using META-R (Alvarado et al., 2020), which employs the LME4 R-package (Bates et al., 2015) for linear mixed model analysis. The Pearson correlations among traits and environments were estimated based on the BLUEs and BLUPs using the “psych” package in the R environment (R Core Team, 2018). The genetic correlations between the five traits were estimated for individual years using the “BMTME” R package (Montesinos-López et al., 2019).

SNP Genotyping

Fresh leaf tissues were collected from each line for DNA isolation using the hexadecyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). Genotyping-by-sequencing (GBS) was performed following double digestion with HF-*PstI* and *MspI* restriction enzymes for library preparation (Poland et al., 2012a). GBS libraries were sequenced using an IonProton sequencer (Thermo Fisher Scientific, Waltham, MA, United States) at the USDA Central Small Grain Genotyping Lab, Manhattan, KS, United States. TASSEL v5.0 was used to call single-nucleotide polymorphisms (SNPs) using the GBS v2.0 discovery pipeline (Bradbury et al., 2007). The reads were aligned to the Chinese Spring wheat genome reference RefSeq v1.1 (IWGSC, 2018) using the default settings of Burrows-Wheeler Aligner v0.6.1. For quality control, SNPs with more than 20% missing data points and minor allele frequency (MAF) of <0.05 were removed. Finally, we obtained 10,290 high-quality SNPs after removing the SNPs that were unmapped on any wheat chromosome. The missing data points in the selected SNP set were imputed using BEAGLE v4.1 (Browning and Browning, 2007). The additive relationship matrix for GP models was estimated using the *A.mat* function in the “rrBLUP” package in R (Endelman, 2011). The Kinship (K)-based marker matrix was estimated using the Centered IBS (identity by state) method (Endelman and Jannink, 2012) implemented through Genomic Association and Prediction Integrated Tool (GAPIT) (Tang et al., 2016).

Genomic Prediction Models and Cross-Validation

We evaluated one univariate and two multivariate GP models for predicting five agronomic traits. Different cross-validation schemes that mimic actual scenarios in a breeding program were used to estimate the PA of these traits and compare the performance of different models.

Single-Trait Model

Ridge regression best linear unbiased prediction (Endelman, 2011) is the commonly used GS model in plant breeding. Similar to the genomic best linear unbiased prediction (GBLUP) model, rrBLUP assumes the normal distribution of marker effects with equal variance. We used rrBLUP as a baseline GS model for all the traits to evaluate the performance of multivariate models. The within-environment trait BLUEs were calculated and then used as input to perform rrBLUP within each environment. A linear

mixed model was implemented using the following model:

$$y = 1\mu + Zu + \varepsilon \quad (4)$$

where y is the vector ($n \times 1$) of adjusted means (BLUEs) from n genotypes for a given trait; μ is the overall mean; Z is the design matrix ($n \times p$) with known values of p markers for n genotypes; u is a genotypic predictor with $u \sim N(0, G_{n \times n} \sigma_g^2)$, where G is positive semi-definite matrix, obtained from markers using “*A.mat*,” which is an additive relation matrix function and σ_g^2 is the additive genetic variance; ε is the residual error with $\varepsilon \sim N(0, \sigma_e^2)$.

Multi-Trait Model

A Bayesian Multivariate Gaussian model with an unstructured variance-covariance matrix was used for the multi-trait model (MT) (Lado et al., 2018). The MT model can be described as

$$y = 1\mu + Zu + \varepsilon \quad (5)$$

where y is the vector with a length of $n \times t$ (n genotypes and t traits); μ is the means vector; Z represents the incidence matrix of order $[(n \times t)p]$; $u_{[(n \times t)p]}$ is a genotypic predictor for all individuals and traits with $u \sim N(0, \sum \otimes G)$. The matrix G represents the positive semi-definite matrix obtained from markers. The residuals of the MT model are represented by the vector ε , with $\varepsilon \sim N(0, R \otimes I)$. The matrices $\sum$ and R are the variance-covariance matrices for depicting the genetic and residual effects, respectively, for each individual in all traits, estimated with the Gibbs sampler with 5,000 burn-in and 25,000 iterations in R package “MTM” (de los Campos and Grüneberg, 2016). The $\sum$ was estimated as an unstructured matrix and R as a diagonal matrix following Lado et al. (2018).

Bayesian Multi-Trait Multi-Environment Model

The Bayesian multi-trait multi-environment model for genomic predictions (Montesinos-López et al., 2016, 2019) can be briefly described as

$$y = X\beta + Z_1b_1 + Z_2b_2 + \varepsilon \quad (6)$$

where y is the response matrix of order $j \times t$ (where t is the number of traits and $j = n \times l$, where n denotes the number of genotypes and l denotes number of environments); X is the design matrix for environmental effects of order $n \times l$, whereas β is the matrix of beta coefficients of order $l \times t$. Z_1 is the incidence matrix of genotypes of order $j \times n$, and b_1 is the matrix of genotypic random effects of order $n \times t$. Z_2 is the incidence matrix of genotype $\times$ environment interaction of order $j \times ln$ and b_2 is the random effect of genotype $\times$ environment $\times$ traits of order $ln \times t$. We assume that b_1 is distributed under a matrix variate normal distribution as $b_1 \sim MN(0, G, \sum_t)$, where G is of order $n \times n$, obtained from SNP markers using “*A.mat*,” which is an additive relation matrix function in rrBLUP, and $\sum_t$ is the unstructured variance-covariance matrix of traits of order $t \times t$. The b_2 is assumed to be distributed under a matrix variate normal distribution as $b_2 \sim MN(0, \sum_E \otimes G, \sum_t)$, where $\otimes$ denotes a Kronecker product and $\sum_E$

is the unstructured variance-covariance matrix of $l \times l$. The matrix ε is the matrix of residuals of order $j \times t$ distributed as $\varepsilon \sim MN(0, I_j, R_e)$. A detailed account of this model and prior distributions can be found in Montesinos-López et al. (2019). Model simulations were carried out using the R package “BMTME” (Montesinos-López et al., 2019) with 5,000 burn-in and 25,000 iterations.

Assessment of Prediction Ability

Predictive ability was estimated as Pearson correlation coefficient between genome-estimated breeding values and observed phenotypes for the testing set of breeding lines. The PA for the rrBLUP model was estimated using cross-validation scheme 1 (CV1), where the population was equally divided into five subpopulations, with four subpopulations (80%) as the training population (phenotyped and genotyped) to train the model and one subpopulation (20%) as the testing population (genotyped only) for prediction. The single-trait model with cross-validation scheme 1 (designated as ST-CV1 hereafter) was implemented in the “rrBLUP” R package (Endelman, 2011) for one trait at a time. The cross-validation process was repeated 1,000 times, and each iteration included different lines in the training and testing sets.

The prediction accuracy (PA) of the MT model was estimated using two cross-validation schemes, as described in Lado et al. (2018) (**Supplementary Figure 1**). Similar to the ST-CV1 scheme, the first cross-validation scheme (MT-CV1) used a random set of lines (80%) as a training set and the remaining lines (20%) as a testing set. The model was trained using genotypic and phenotypic data of these lines in the training set, and only genotypic data were used to predict the performance of the testing set lines based on the model built from the training set. This process of splitting the data into training and testing sets was repeated 50 times. Hence, a different set of lines were selected into the training and testing datasets for each iteration. The CV1 scheme mocks the breeding situation where a set of lines that are evaluated for given traits could be used to predict an unphenotyped set of lines that only have genotypic information. In the second cross-validation scheme (MT-CV2), the lines were randomly split into a training set (80%) and a testing set (20%). To train the model, MT-CV2 used genotypic data and phenotypic data of secondary traits from both the training and testing sets, but the phenotypic data of the target trait (primary trait) only from the training set. The BMTME model used a cross-validation scheme similar to MT-CV1 to estimate the PA of the model by randomly splitting the lines into an 80% training set and a 20% testing set. Since the BMTME model employs a Gibbs sampler with multiple iterations and is computationally expensive, the cross-validation scheme was repeated only 25 times.

Application of MTME Genomic Prediction in the Breeding Program

As the multi-trait multi-environment model showed a potential in predicting different agronomic traits the using cross-validation approach, we evaluated the possible application of this method in the breeding program to reduce phenotyping efforts and per plot costs. As discussed earlier, we evaluated ~40 elite lines and ~110 advanced lines each year in multiple environments. Per

plot costs and phenotyping efforts could be reduced if we can successfully determine the genomic estimation of breeding values (GEBVs) of the advanced lines in fewer locations rather than testing these lines in all available locations. The MTME model can estimate the environmental effect based on elite lines evaluated in all locations and the genotypic effect of advanced lines from fewer locations. To test this, we used the MTME model in an allocation design where we used the phenotypic data of elite lines from five testing environments; however, we used phenotypic records of advanced lines from three environments only. We predicted five traits in the remaining two environments in both the growing seasons. The model was fitted using the R package “BMTME” (Montesinos-López et al., 2016, 2019) with 5,000 burn-in and 15,000 iterations. The observed phenotypic records from the remaining two environments were used to assess the predictive accuracy of the design.

RESULTS

Descriptive Statistics

The phenotypic BLUEs for grain yield, grain protein content, test weight, plant height, and days to heading varied significantly among the different environments (**Table 1**). HYS produced the highest mean grain yield in both years, where BRK and WIN produced the lowest grain yield in 2018–19 and 2019–20, respectively. Broad-sense heritability (H^2) was estimated for all the five agronomic traits in each environment (**Table 1**). Differences in heritability estimates (0.63–0.96) describe the different genetic architecture of traits and contrasting environmental effects. Among the five traits evaluated in the study, TW, HT, and HDs had moderate to high heritability values in most of the environments and over both years. Relatively, YLD (0.64–0.84) and PROT (0.63–0.96) had comparatively lower heritability than other traits. Among the five environments, the heritability for all the traits was high in both the experimental years in DL. For YLD heritability, HYS (2019–20) had the highest (0.84), whereas BRK (2019–20) had the lowest (**Table 1**).

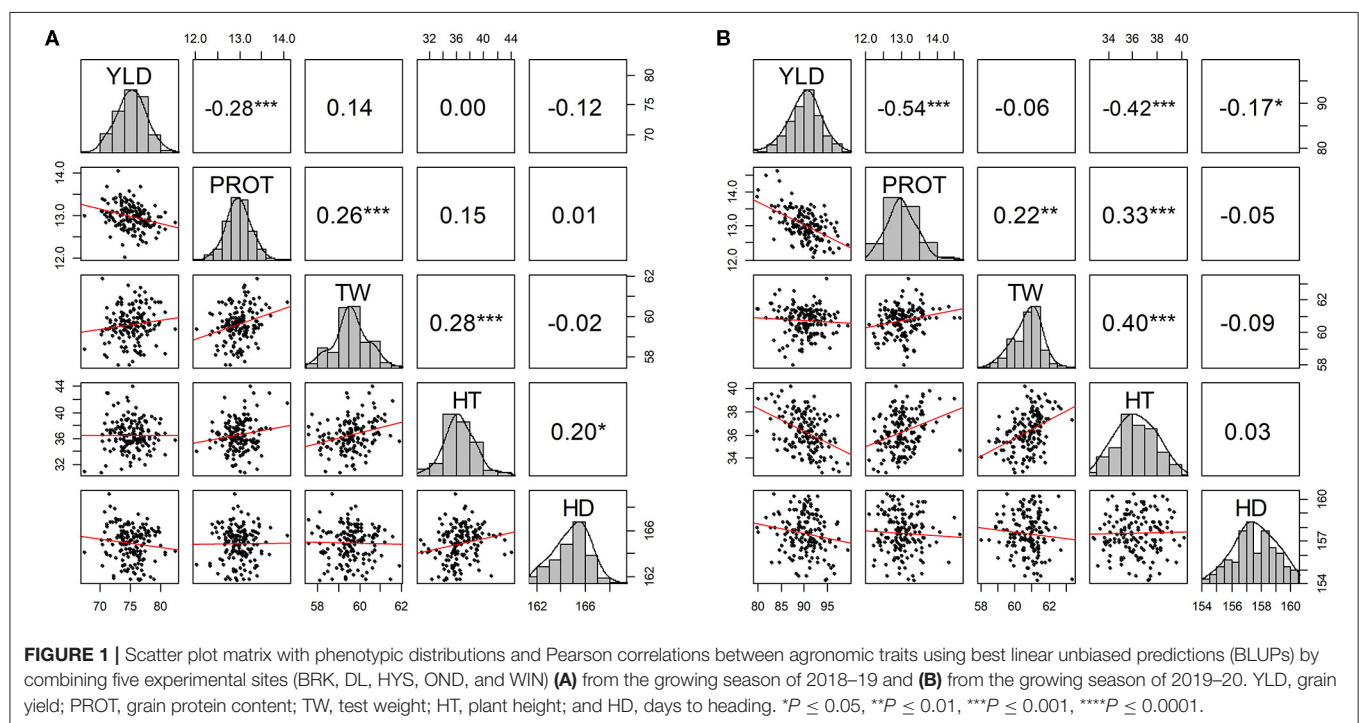
Pearson correlations among agronomic traits were calculated using BLUEs by combining phenotypic data from all environments in each of the two growing seasons (**Figure 1**). As expected, significant negative correlation values (–0.28 and –0.54) were observed between YLD and PROT in both years. YLD was also negatively correlated with HDs (in both years) and HT (2019–20) (**Figure 1**). Similarly, TW was positively correlated with PROT and HT in both growing seasons. Overall, higher correlation values were observed between the agronomic traits in the 2019–20 growing season than in 2018–19 (**Supplementary Figures 2, 3**). Furthermore, genetic correlations among the five traits are estimated by fitting the BMTME model for individual growing seasons and are presented in **Supplementary Tables 2, 3**. Similar to the phenotypic correlation estimates, we observed a higher genetic correlation in 2019–20 as compared to 2018–19.

We further estimated the Pearson correlations among the five environments in 2018–19 and 2019–20 using the data of all the five agronomic traits (**Supplementary Figure 4**).

TABLE 1 | Trait descriptive statistics and broad-sense heritability estimate for individual site-year environments of lines grown over five locations (Env) in 2018–19 and 2019–20 growing seasons.

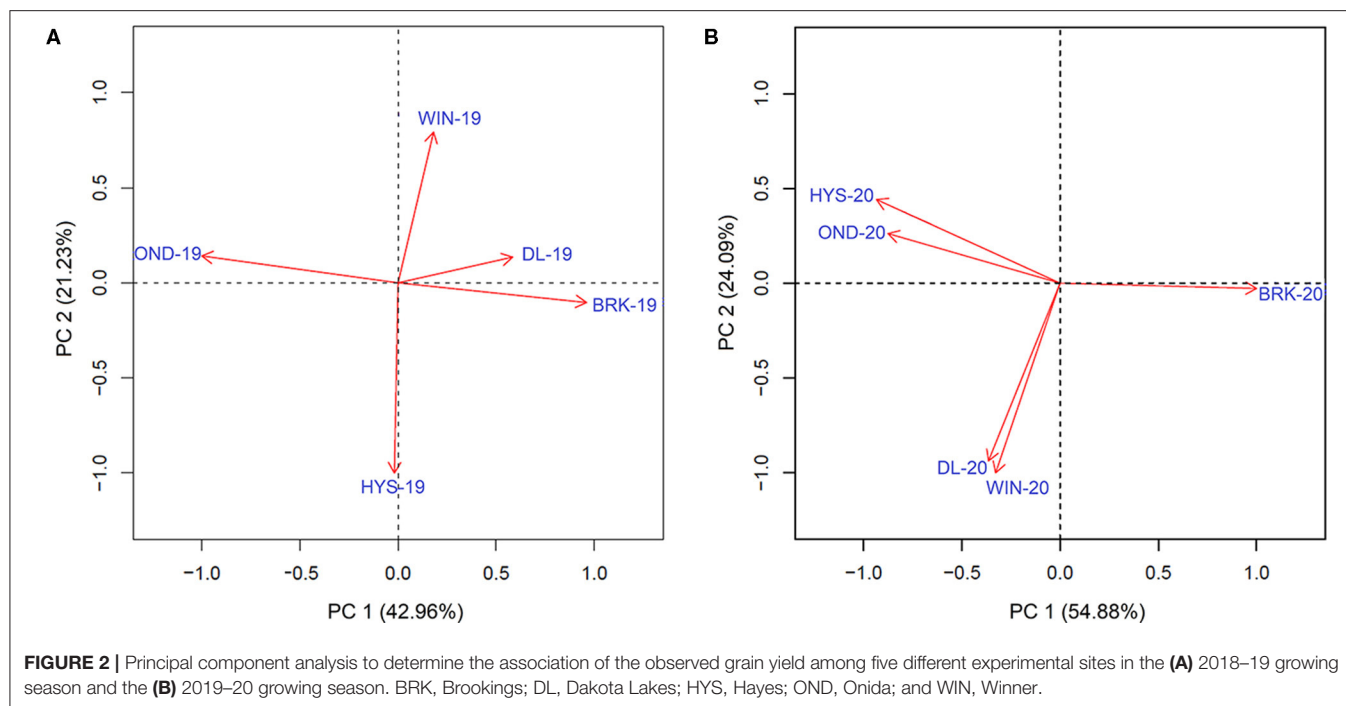
Year	Env ^a	Yield (bu ac ⁻¹)			Protein content (%)			Test weight (kg hL ⁻¹)			Plant height (cm)			Days to heading (julian days)		
		GM ^b	CV	H ²	GM	CV	H ²	GM	CV	H ²	GM	CV	H ²	GM	CV	H ²
2018-19	BRK	64.69	8.96	0.80	12.16	5.45	0.69	72.51	1.42	0.91	94.78	4.25	0.89	163.22	0.48	0.92
	DL	77.44	6.48	0.77	14.25	1.55	0.94	77.09	1.35	0.78	86.06	3.99	0.89	164.65	0.76	0.92
	HYS	81.98	6.46	0.73	12.04	3.66	0.72	76.55	1.06	0.92	99.12	3.18	0.90	163.84	0.70	0.74
	OND	71.21	7.27	0.76	13.25	3.30	0.85	78.21	1.27	0.82	89.91	3.62	0.91	168.73	0.65	0.87
	WIN	81.27	5.89	0.79	13.17	4.27	0.63	79.12	1.00	0.88	93.63	2.90	0.95	164.19	0.80	0.89
2019-20	BRK	84.26	6.25	0.64	12.49	3.65	0.80	77.39	0.90	0.89	86.67	4.42	0.75	156.18	0.63	0.89
	DL	93.31	4.14	0.78	13.55	1.40	0.96	79.10	0.64	0.95	85.80	3.45	0.83	155.74	0.40	0.94
	HYS	96.64	4.66	0.84	13.87	2.02	0.90	77.26	1.30	0.91	102.8	3.95	0.82	159.36	0.51	0.85
	OND	92.21	4.40	0.81	11.99	4.97	0.59	78.65	1.08	0.87	92.53	3.59	0.85	157.09	0.63	0.87
	WIN	84.16	4.73	0.80	13.24	2.99	0.84	78.24	0.90	0.89	92.70	3.47	0.85	158.75	0.63	0.91

BRK, Brookings; DL, Dakota Lakes; HYS, Hayes; OND, Onida; and WIN, Winner.

^aEnv, refers to different trial location. BRK, Brookings; DL, Dakota Lakes; HYS, Hayes; OND, Onida; and WIN, Winner.^bGM, general mean for respective trait; CV, coefficient of variation; H², broad sense heritability.

Significantly higher correlation values were observed for YLD among the five environments in 2019–20 than those in 2018–19. A similar trend was observed for PROT, TW, and HDs; however, correlations were comparable for HT between the two growing seasons (Supplementary Figure 4). Moreover, the principal component analysis (PCA) on YLD validated strong correlations among the testing locations, in particular

between HYS and OND and between DL and WIN, in the 2019–20 growing season (Figure 2). However, only a weak correlation was observed between DL and BRK in the 2018–19 growing season. The varying degrees of correlation among the locations in different growing seasons provide an opportunity to compare the performance of the MTME model in different growing environments.



Genetic Relationship Among Lines

The kinship-based marker relationship matrix was derived using 10,290 SNPs from 151 lines evaluated in the 2018–19 growing season and 156 lines evaluated in the 2019–20 growing season (Supplementary Figure 5). The positive values of the relationship matrix signify an increased likelihood of the allele from one line being detected in other lines. The heatmaps of both the relationship matrices elucidate several small groups of closely related individuals over both the growing seasons. Most of the lines seem genetically related to several others. However, the heatmaps did not reveal any large genetically structured sub-populations in either set of 151 or 156 lines, respectively. Thus, the absence of a strong structure suggests no advantage of performing stratified sampling for cross-validation schemes to estimate PA. Furthermore, the density of heatmaps revealed a closer relationship among the 156 lines evaluated in 2019–20 (Supplementary Figure 5A) than among the 151 lines evaluated in 2018–19 (Supplementary Figure 5B).

Genomic Prediction Using 2018–19 and 2019–20 Datasets

We compared the predicted performance of five traits among four different approaches using two data sets (2018–19 and 2019–20). The PA of various models for the five traits is presented in Supplementary Tables 4, 5. ST-CV1 was used as a baseline model to compare the performance of different multivariate models. In 2018–19, the mean PA using ST-CV1 was 0.31, 0.35, 0.36, 0.35, and 0.36 for YLD, PROT, TW, HT, and HDs (Figure 3). Slightly better performance was observed in 2019–20 where ST-CV1 yielded an average PA of 0.36, 0.35, 0.54, 0.33, and 0.35 for these traits, respectively. The multi-trait model

was tested using two prediction scenarios, MT-CV1 and MT-CV2. The MT-CV1 model did not show improvement in the PA over ST-CV1 for any of the five traits in either growing season (Supplementary Tables 4, 5).

The multi-trait model MT-CV2, which includes phenotypic data for secondary agronomic traits from individuals to be predicted, showed an overall higher prediction accuracy for YLD in both growing seasons. In 2018–19, the PA for YLD using the MT-CV2 model ranged from 0.15 to 0.56, outperforming the single-trait (ST-CV1) model by an average of 26% (Supplementary Tables 4, 5). Similarly, the mean PA for YLD in 2019–20 using MT-CV2 was 0.59, showing 63% improvement over the ST-CV1 model. The best PA for YLD in 2019–20 was observed in HYS (0.71), followed by WIN (0.67) and DL (0.57). The improvement in PA over ST-CV1 reached up to 148% in WIN and 80% in BRK in 2019–20.

Likewise, we observed a marginal to moderate improvement in PA for other agronomic traits using MT-CV2 model in both of the growing seasons (Figures 3, 4 and Supplementary Tables 4, 5). In 2018–19, the mean PA using MT-CV2 was 0.4, 0.42, 0.34, and 0.38 for PROT, TW, HT, and HDs, exhibiting an improvement of 14, 19, 36, and 8%, respectively. In comparison, the PA using MT-CV2 was higher in 2019–20, with an average PA of 0.54, 0.59, 0.43, and 0.38 for PROT, TW, HT, and HDs with an improvement of 54, 9, 30, and 8%, respectively. Overall, the better performance of the MT-CV2 model can be attributed to the higher genetic correlation among the traits evaluated in 2019–20 over the 2018–19 season (Supplementary Tables 2, 3).

The multi-trait multi-environment MTME model generalizes the multi-trait model to consider the correlation among the environments on top of the genetic correlation between the

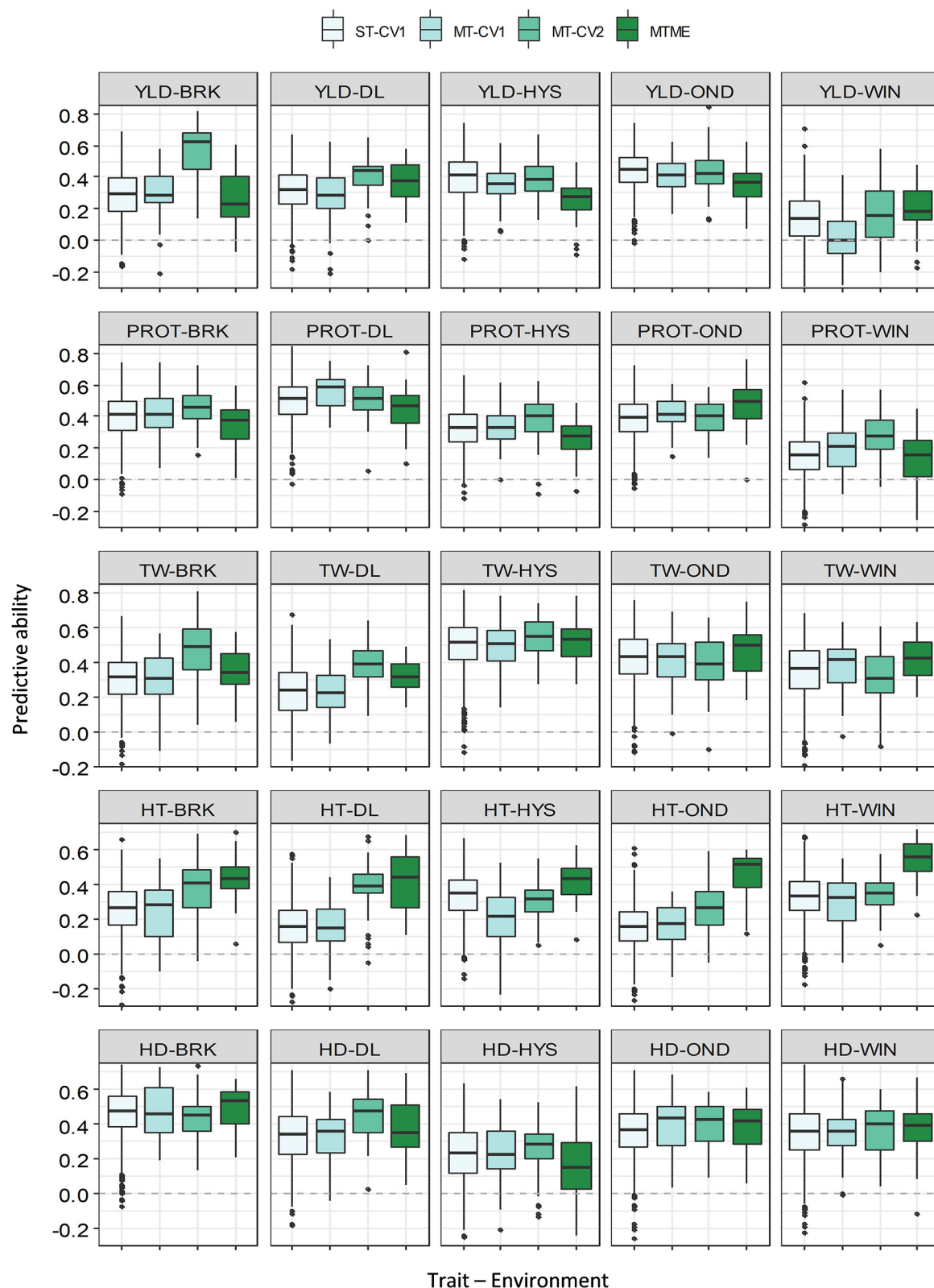


FIGURE 3 | Prediction accuracy (PA) for five agronomic traits evaluated in five environments in the growing season of 2018–19. Boxplots compare the PA using a single-trait prediction model with one cross-validation scheme (ST-CV1), a multi-trait prediction model with two cross-validation schemes (MT-CV1 and MT-CV2), and a Bayesian multi-trait multi-environment prediction model (MTME). Traits include YLD, grain yield; PROT, grain protein content; TW, test weight; HT, plant height; and HD, days to heading.

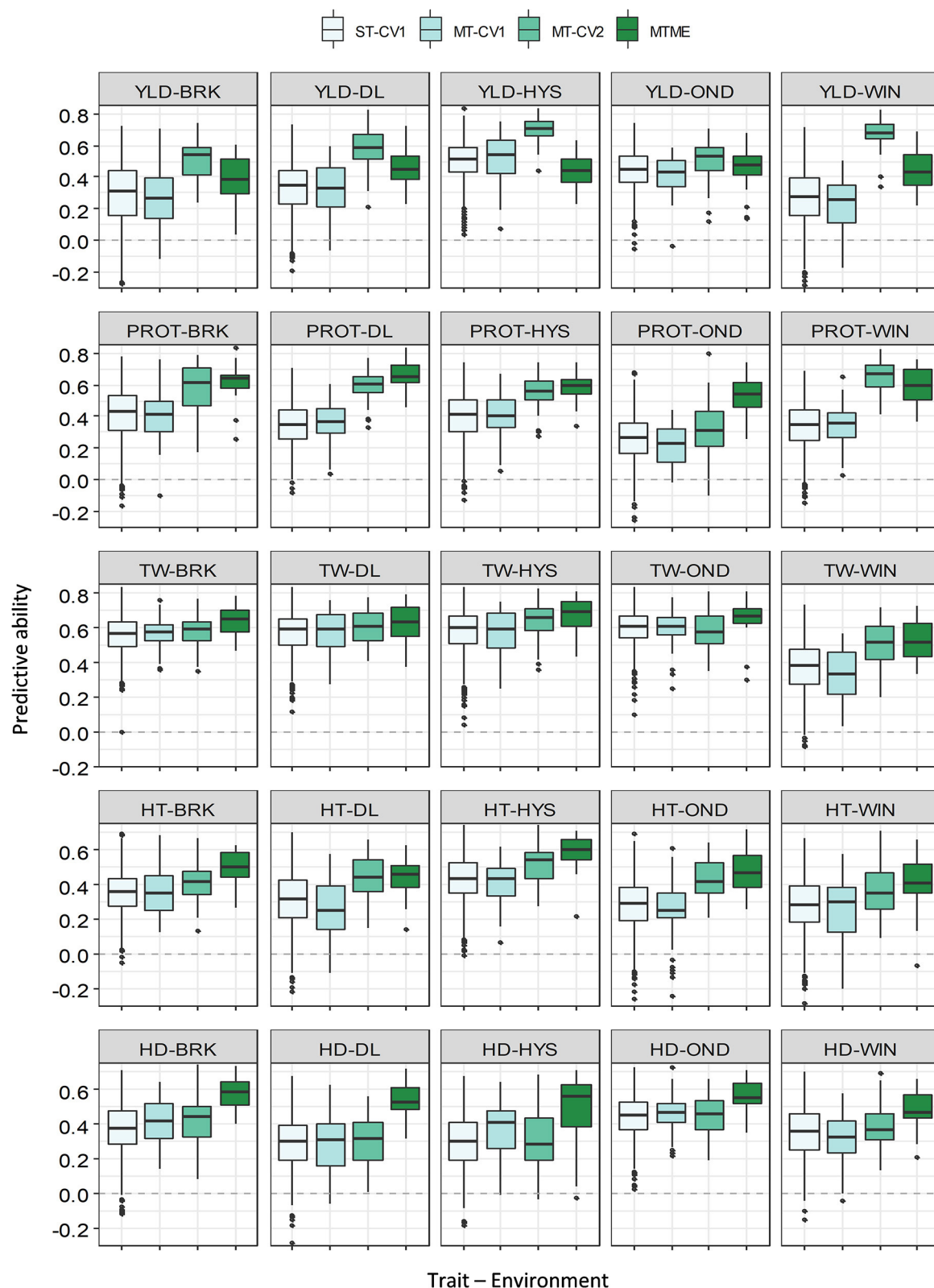


FIGURE 4 | Prediction accuracy (PA) for five agronomic traits evaluated in five environments in the growing season of 2019–20. Boxplots compare the PA using a single-trait prediction model with one cross-validation scheme (ST-CV1), a multi-trait prediction model with two cross-validation schemes (MT-CV1 and MT-CV2), and a Bayesian multi-trait multi-environment prediction model (MTME). Traits include YLD, grain yield; PROT, grain protein content; TW, test weight; HT, plant height; and HD, days to heading.

traits. In 2018–19, the MTME model did not show a significantly different PA over the ST-CV1 model for YLD (0.18–0.36) and PROT (0.13–0.46). The performance of the MTME model for these two traits likely relates to the lower genetic trait correlations and lower correlation among the environments for these traits in 2018–19 (**Supplementary Figure 4**). Analogous to YLD and PROT, the MTME model resulted in a higher prediction accuracy than the ST-CV1 model for TW, HT, and HDs in 2018–19 (**Figure 3**). For instance, the average PA using MTME for TW, HT, and HDs was 0.42, 0.42, and 0.36, respectively, which translates to an improvement of 19, 68, and 12%, respectively. Furthermore, the PA using the MTME model outstripped the ST-CV1 model in all the five environments for TW (0.32–0.52) and HT (0.41–0.54), and in four environments for HDs (**Figure 3**).

In contrast to 2018–19, we observed higher genetic correlations among the five traits and higher environmental correlations in 2019–20 (**Supplementary Tables 2, 3** and **Supplementary Figure 4**). As a result of high correlation values, we observed a consistent improvement in the PA of MTME in all the environments for all the five traits (**Figure 4** and **Supplementary Table 4**). For YLD, the MTME model also performed better than the single-trait model in most of the environments, except HYS. The average PA for YLD using the MTME model was 0.43, which was 22% better than the ST-CV1 model. Furthermore, the MTME model appeared to be superior for predicting PROT and TW (**Figure 4**). For PROT, the MTME model performed best in all the locations, with a PA ranging from 0.52 to 0.67 (**Supplementary Table 5**). We achieved an improvement in PA of up to 100% (OND) using the MTME model (0.52) over the single-trait model (0.26) with 71% improvement on average. The PA for TW was higher using the MTME model than the other models, ranging from 0.53 to 0.67, with a mean improvement of 17% over the ST-CV1 model (**Supplementary Table 5**). Similarly, the average PA of the MTME model was the highest for HT (0.49) and HDs (0.53), which outstrips the ST-CV1 model by 48 and 51%, respectively.

Application of MTME Model in the Breeding Program

Based on the cross-validation results, we evaluated the efficacy of the MTME model in reducing phenotypic efforts in the breeding program. We used the MTME model to estimate the GEBV of advanced lines in environments where only elite lines are evaluated. In the tested allocation design, we used phenotypic data of EYTs from five environments and AYT lines from three environments to predict GEBVs of AYT lines in remaining environments (**Figure 5**). Two environments, OND and WIN, were used as testing environments for predicting AYT lines. For 2018–19, we predicted the performance of 96 AYT lines, whereas 2019–20 comprised a prediction of 114 AYT lines in two environments. **Table 2** elucidates the predictive ability for the five agronomic traits using MTME in an independent prediction scenario. Moderate PA was observed for all the traits in both environments except for WIN in 2019–20. For OND, the results showed

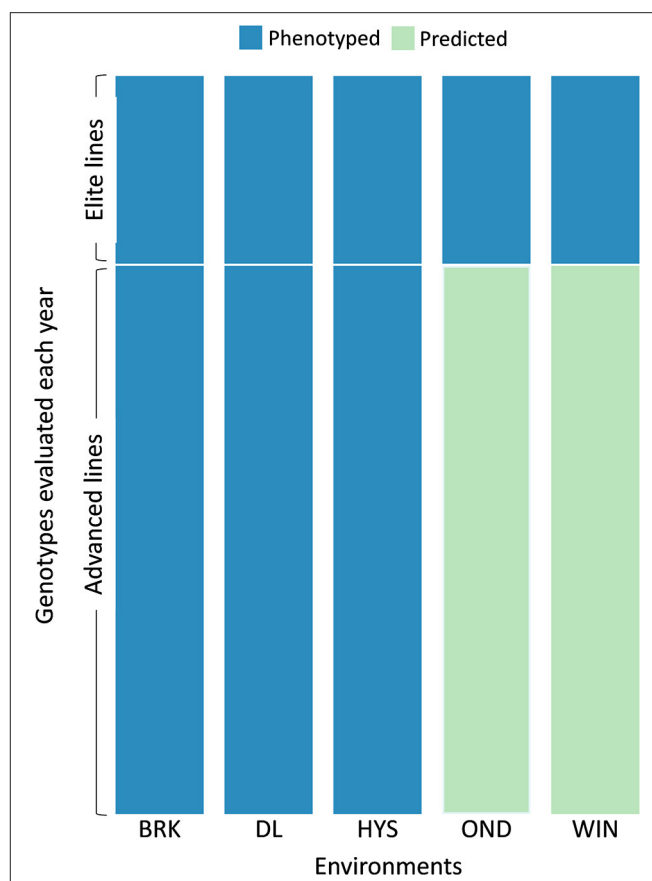


FIGURE 5 | Testing design for the independent prediction of agronomic traits using the MTME model. Each year, a set of elite and advanced lines is evaluated over multiple locations. The sparse testing design proposes phenotyping of elite lines in all the environments (five in this scenario) and advanced lines in fewer environments (three in this scenario). For independent prediction, the dataset from 2018–19 comprised 55 elite lines with checks and 96 advanced lines. The 2019–20 dataset comprised 42 elite lines with checks and 114 advanced lines. Five environments: BRK, Brookings; DL, Dakota Lakes; HYS, Hayes; OND, Onida; and WIN, Winner.

a better prediction accuracy than WIN for YLD and TW. Overall, the results suggest that the MTME model could be used by evaluating an overlapping set of lines over multiple environments and lines in early testing could be tested in fewer environments.

DISCUSSION

In recent years, genomic prediction has been intensively evaluated in wheat breeding programs to select and advance lines for several traits of interest (Rutkoski et al., 2014, 2016; Haile et al., 2020; Juliana et al., 2020). However, improving the PA of complex traits remains a challenge for successfully implementing GS in breeding programs. The choice and optimization of statistical models are crucial to improve the performance of GS. Most plant breeding programs currently rely on univariate genomic prediction models to target a single trait at a time. An

TABLE 2 | Predictive ability for the independent prediction of advanced breeding lines (AYTs) in new environments using the MTME model.

Year	Env ^a	Predictive ability ^b				
		Grain yield	Grain protein	Test weight	Plant height	Days to heading
2018–19	OND	0.44	0.37	0.43	0.49	0.27
	WIN	0.30	0.25	0.38	0.30	0.46
2019–20	OND	0.36	0.27	0.44	0.22	0.41
	WIN	0.15	0.32	0.25	0.18	0.24

Tables shows Pearson correlation between the observed and predictive values of agronomic traits in the AYT in two different environments over two growing seasons.

^aEnv refers to different trial location. BRK, Brookings; DL, Dakota Lakes; HYS, Hayes; OND, Onida; and WIN, Winner.

^bThe predictive ability for five agronomic traits using the MTME model in independent prediction of advanced lines. Refer to **Figure 5** for the design of the prediction scheme.

advantage of multivariate prediction approaches over single-trait models that have been demonstrated in some recent studies is utilizing correlations between multiple traits and environments (Jia and Jannink, 2012; Sun et al., 2017; Lado et al., 2018; Ward et al., 2019; Ibba et al., 2020). This study evaluated the application of multi-trait and multi-environment prediction models to predict five key traits of varying genetic architecture across diverse environments in a breeding program.

The ridge-regression best linear unbiased prediction (rrBLUP) is one of the most often used single-trait prediction models. The rrBLUP has an advantage over Bayesian models in predicting complex traits governed by several loci with small effects (Lorenz et al., 2011). We used rrBLUP as a baseline model (ST-CV1) for comparison with different multivariate approaches. The PA for agronomic traits using ST-CV1 was comparable with other studies using the same model (Pérez-Rodríguez et al., 2012; Charmet et al., 2014; He et al., 2016; Maulana et al., 2021). For instance, the PA for YLD was between 0.13 and 0.43 for 2018–19 and 0.27 and 0.5 for 2019–20. The PA for TW in both growing seasons was higher than the PA for other traits because of the highly heritable nature of this trait (**Figures 3, 4**).

We evaluated the multi-trait model using two cross-validation schemes. The first scheme (MT-CV1) conducts multi-trait prediction for new un-phenotyped individuals, and the testing set has not been phenotyped for any of the traits. In the second cross-validation scheme (MT-CV2), phenotype information for the predicted trait is missing, whereas phenotype information for the secondary traits is available in the testing set (Lado et al., 2018; Bhatta et al., 2020). In this study, the PA of the MT-CV1 model was found similar to that of the ST-CV1 model for most of the trait-environment combinations in both growing seasons (**Supplementary Tables 4, 5**). Several studies have reported marginal or no improvement with MT-CV1, where information from secondary traits is limited to the training set (Calus and Veerkamp, 2011; Lado et al., 2018; Schulthess et al., 2018; Arojju et al., 2020; Bhatta et al., 2020). However, other studies reported an improvement in GP when the MT-CV1 model included secondary traits with moderate-high heritability (Jia and Jannink, 2012; Rutkoski et al., 2012; Guo et al., 2014). Jia and Jannink, 2012 suggested that the MT-CV1 approach might be more useful when the primary trait has very low heritability ($H^2 < 0.2$). In this study, the similarity in performance of the MT-CV1 and ST-CV1 models might be contributed by the moderate

to high heritability estimated for most of the traits and the small size of the training population.

In contrast to MT-CV1, the MT-CV2 model significantly improved the PA for all agronomic traits in all the environments, suggesting that the inclusion of secondary traits in the training and testing sets improves the predictive performance of complex traits (**Supplementary Tables 4, 5**). Several studies have reported a similar improvement in prediction using the MT-CV2 model for agronomic and end-use quality traits in wheat (Rutkoski et al., 2016; Sun et al., 2017; Lado et al., 2018), rice (Wang et al., 2017), barley (Bhatta et al., 2020), sorghum (Fernandes et al., 2018), and ryegrass (Arojju et al., 2020). The MT-CV2 model outperformed the single-trait model for YLD prediction in all environments. However, the extent of improvement using the MT-CV2 model varied with traits and environments tested. As multi-trait models rely on the genetic correlation between traits (Calus and Veerkamp, 2011; Jia and Jannink, 2012), differences in prediction improvements due to the MT-CV2 model can be attributed to the varying degrees of genetic correlations observed in different environments. We observed a high genetic correlation among the traits in 2019–20 that resulted in a higher prediction accuracy for the different traits in this growing season (**Figure 1** and **Supplementary Tables 2, 3**). The results suggest that MT-CV2 could likely be very useful if we can include data for HT, HDs, and other spectral indices recorded using a high throughput method for predicting YLD. In addition, the MT-CV2 approach could be really useful to predict hard-to-phenotype end-use quality traits by the inclusion of already available agronomic data for the testing set.

We also evaluated the BMTME model (referred to as MTME) that generalizes a multi-trait model to consider the correlations among multiple environments. Recently, two studies reported an increase in the PA of agronomic and end-use quality traits in wheat using the BMTME approach (Guo et al., 2020; Ibba et al., 2020). Because of the different training process, we did not directly compare the MTME model with the MT-CV2 model but compared both with the ST-CV1 model. In 2018–19, the MTME model proved to be better than the ST-CV1 and MT-CV1 models for all the traits except YLD and PROT. However, the MTME model outperformed the ST-CV1 and MT-CV1 models in 2019–20 for all the traits in all the environments (**Supplementary Table 5**). The mean improvement in PA (across five environments) using MTME

model over the ST-CV1 reached up to 19, 71, 17, 48, and 51% for YLD, PROT, TW, HT, and HDs, respectively. The differences in performance of the MTME model in 2019–20 compared with 2018–19 relate to the observed genetic correlations among the traits as well as among the environments in these growing seasons (**Supplementary Figure 2A**). As discussed earlier, the genetic correlations between traits and correlation among environments were higher in 2019–20 compared with those in 2018–19. Thus, a higher PA was observed for the traits showing a high correlation among the different environments. For example, the five environments were highly correlated for PROT (0.56–0.76) compared with YLD (0.23–0.65) (**Supplementary Figures 3, 4**), explaining the difference in the improvement of PA for these traits. Overall, the results suggest that the MTME model could be successfully applied in a program if there is a moderate to high correlation for a trait between environments and overcome the effect of a small training population.

Apart from the statistical model, the heritability (H^2) of a trait is another crucial factor for improving PA (Lorenz et al., 2011; Combs and Bernardo, 2013). Several studies have found that low heritability often results in lower prediction accuracy in single-trait genomic prediction (Heffner et al., 2009; Jannink et al., 2010). The application of multi-trait models can improve the PA of low-heritability traits using the information from correlated traits with high heritability (Jia and Jannink, 2012; Jiang et al., 2015; Lado et al., 2018; Bhatta et al., 2020). The heritability estimates for most of the traits in different environments were moderate to high in this study, with few exceptions. The use of the MT-CV2 model significantly improved the predictive ability for PROT in WIN (0.15 to 0.29) and TW in DL (0.23 to 0.39), where highly heritable and moderately correlated traits were included in the model. In contrast, the MT-CV2 model did not improve the PA for HDs in HYS (0.23 to 0.25), as the primary trait was weakly correlated to the highly heritable secondary traits in the model. The results suggest that the inclusion of highly heritable but weakly correlated secondary traits in the multi-trait model may not improve the PA.

Genomic prediction has been suggested to implement sparse testing in multi-environment trials and reduce the resources involved in phenotyping (Jarquin et al., 2020). Based on the promising cross-validation results using MTME models, we evaluated the application of this model in the breeding program to reduce phenotyping resources. At the SDSU winter wheat breeding program, we evaluate a set of elite (EYTs) and advanced (AYTs) lines each year in multiple environments. However, the results suggest that GP models developed using phenotypic data from all locations of EYTs and limited locations of AYT lines can predict AYT lines in remaining environments (**Table 2**). This strategy could be useful as we evaluate ~40 EYTs and ~110 AYT lines each year in replicated nurseries and testing the AYT plots at two/three locations instead of five can save substantial resources. Though we used this strategy to predict AYT lines at two locations, further improved GP models assisted with enviro-omics data can help to predict more environments with better accuracy. Moreover, this strategy can be expanded to predict preliminary breeding lines at earlier testing stages.

In conclusion, this study evaluated the PA of univariate and multivariate GP models for five agronomic traits in advanced winter wheat breeding lines. We compared two different cross-validation strategies mocking practical breeding scenarios. Overall, the results supported the practical implementation of multivariate GS models in predicting complex traits. We found a significant advantage of using MT and MTME models when correlated traits and/or environments are included in the models. The results suggest that the inclusion of correlated traits and environments in prediction models can offset the limitation of a small training population, allowing the use of advanced breeding lines to predict preliminary breeding lines in the same year or the following one. It will be interesting to further study the inclusion of different combinations of secondary traits in the MT model to increase the PA of YLD. We envision that the evaluation of secondary traits such as plant height, tillers/m², spike length, and spike density that have high correlations with YLD using an unmanned aerial system (UAS) in winter wheat yield trials could help predict YLD. This would permit trials on a large number of locations (e.g., >10) but harvesting only in a limited number (e.g., 2–3) of locations. Similarly, evaluating secondary traits (grain protein, flour protein, water absorption, gluten content, and quality) could facilitate the prediction of other complex traits such as end-use quality. Finally, GS holds a tremendous potential for improving the selection accuracy of complex traits in wheat breeding; however, we believe GEBVs will complement phenotyping efforts rather than replacing them. Future breeding strategies should focus on increasing the efficiency of breeding programs by maximizing genetic gain.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

SS and HG conceptualized the experiment and designed the methodology. SS, HG, CH, NB, JH, and JZ performed the field trials. HG and SS performed the data curation, performed the data analysis, software implementation, and visualization, and wrote the original manuscript. PA, AB, and GB performed genotyping and SNP discovery. TR assisted in data analysis and software implementation. GB, PA, EO, BT, JH, and SA contributed to the interpretation of results and revision of the manuscript. SS obtained the funding for the project. All authors approved the manuscript.

FUNDING

This project was collectively funded by the USDA hatch projects SD00H538-15 and SD00H695-20 and the Agriculture and Food Research Initiative Competitive Grants 2017-67007-25939 (Wheat-CAP), USDA-ARS agreement 59-0206-0-177 and 2019-67013-29015 from the USDA National Institute of Food and

Agriculture and South Dakota Wheat Commission grant 3X0281. The funders had no role in the study design, data collection, analysis, decision to publish, or manuscript preparation.

ACKNOWLEDGMENTS

The authors would like to thank the South Dakota Agriculture Experimental Station (Brookings, SD, United States) for providing the resources to conduct the experiments. The mention of trade names or commercial products in this publication is

solely for the purpose of providing specific information and does not imply recommendation or endorsement by the United States Department of Agriculture. The USDA is an equal opportunity provider and employer.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.709545/full#supplementary-material>

REFERENCES

- Alvarado, G., Rodríguez, F. M., Pacheco, A., Burgueño, J., Crossa, J., Vargas, M., et al. (2020). META-R: a software to analyze data from multi-environment plant breeding trials. *Crop J.* 8, 745–756. doi: 10.1016/j.cj.2020.03.010
- Arojiu, S. K., Cao, M., Trolove, M., Barrett, B. A., Inch, C., Eady, C., et al. (2020). Multi-trait genomic prediction improves predictive ability for dry matter yield and water-soluble carbohydrates in perennial ryegrass. *Front. Plant Sci.* 11:1. doi: 10.3389/fpls.2020.01197
- Arruda, M. P., Brown, P. J., Lipka, A. E., Krill, A. M., Thurber, C., and Kolb, F. L. (2015). Genomic selection for predicting *Fusarium* head blight resistance in a wheat breeding program. *Plant Genome* 8, 1–12. doi: 10.3835/plantgenome2015.01.0003
- Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., and Crossa, J. (2015). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Sci.* 242, 23–36. doi: 10.1016/j.plantsci.2015.08.021
- Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Battenfield, S. D., Guzmán, C., Gaynor, R. C., Singh, R. P., Peña, R. J., Dreisigacker, S., et al. (2016). Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. *Plant Genome* 9, 1–12. doi: 10.3835/plantgenome2016.01.0005
- Belamkar, V., Guttieri, M. J., Hussain, W., Jarquín, D., El-basyoni, I., Poland, J., et al. (2018). Genomic selection in preliminary yield trials in a winter wheat breeding program. *G3 Genes Genomes Genet.* 8, 2735–2747. doi: 10.1534/g3.118.200415
- Bhat, J. A., Ali, S., Salgotra, R. K., Mir, Z. A., Dutta, S., Jadon, V., et al. (2016). Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front. Genet.* 7:221. doi: 10.3389/fgene.2016.00221
- Bhatta, M., Gutierrez, L., Cammarota, L., Cardozo, F., Germán, S., Gómez-Guerrero, B., et al. (2020). Multi-trait genomic prediction model increased the predictive ability for agronomic and malting quality traits in barley (*Hordeum vulgare* L.). *G3 Genes Genomes Genet.* 10, 1113–1124. doi: 10.1534/g3.119.400968
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Browning, S. R., and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084–1097. doi: 10.1086/521987
- Calus, M. P. L., and Veerkamp, R. F. (2011). Accuracy of multi-trait genomic selection using different methods. *Genet. Sel. Evol.* 43:26. doi: 10.1186/1297-9686-43-26
- Charmet, G., Storlie, E., Oury, F. X., Laurent, V., Beghin, D., Chevarin, L., et al. (2014). Genome-wide prediction of three important traits in bread wheat. *Mol. Breed.* 34, 1843–1852. doi: 10.1007/s11032-014-0143-y
- Combs, E., and Bernardo, R. (2013). Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. *Plant Genome* 6, 1–7. doi: 10.3835/plantgenome2012.11.0030
- de los Campos, G., and Grüneberg, A. (2016). *MTM Package*. Available online at: <https://github.com/QuantGen/MTM/>
- de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., and Calus, M. P. L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193, 327–345. doi: 10.1534/genetics.112.143313
- Dong, H., Wang, R., Yuan, Y., Anderson, J., Pumphrey, M., Zhang, Z., et al. (2018). Evaluation of the potential for genomic selection to improve spring wheat resistance to *Fusarium* head blight in the Pacific Northwest. *Front. Plant Sci.* 9:911. doi: 10.3389/fpls.2018.00911
- Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024
- Endelman, J. B., and Jannink, J.-L. (2012). Shrinkage estimation of the realized relationship matrix. *G3 Genes Genomes Genet.* 2, 1405–1413. doi: 10.1534/g3.112.004259
- Fernandes, S. B., Dias, K. O. G., Ferreira, D. F., and Brown, P. J. (2018). Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. *Theor. Appl. Genet.* 131, 747–755. doi: 10.1007/s00122-017-3033-y
- Fischer, R., Byerlee, D., and Edmeades, G. (2014). *Crop Yields and Global Food Security*. ACIAR: Canberra, ACT. Available at: http://www.academia.edu/download/35887178/Crop_yields_and_global_food_security__a_book_by_T.Fischer_et_al_2014.pdf (accessed January 12, 2021).
- Guo, G., Zhao, F., Wang, Y., Zhang, Y., Du, L., and Su, G. (2014). Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet.* 15:30. doi: 10.1186/1471-2156-15-30
- Guo, J., Khan, J., Pradhan, S., Shahi, D., Khan, N., Avci, M., et al. (2020). Multi-trait genomic prediction of yield-related traits in US soft wheat under variable water regimes. *Genes* 11:1270. doi: 10.3390/genes11111270
- Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186. doi: 10.1186/1471-2105-12-186
- Haile, T. A., Walkowiak, S., N'Diaye, A., Clarke, J. M., Hucl, P. J., Cuthbert, R. D., et al. (2020). Genomic prediction of agronomic traits in wheat using different models and cross-validation designs. *Theor. Appl. Genet.* 1:3. doi: 10.1007/s00122-020-03703-z
- Hayes, B. J., Panozzo, J., Walker, C. K., Choy, A. L., Kant, S., Wong, D., et al. (2017). Accelerating wheat breeding for end-use quality with multi-trait genomic predictions incorporating near infrared and nuclear magnetic resonance-derived phenotypes. *Theor. Appl. Genet.* 130, 2505–2519. doi: 10.1007/s00122-017-2972-7
- He, S., Schulthess, A. W., Mirdita, V., Zhao, Y., Korzun, V., Bothe, R., et al. (2016). Genomic selection in a commercial winter wheat population. *Theor. Appl. Genet.* 129, 641–651. doi: 10.1007/s00122-015-2655-1
- Heffner, E. L., Sorrells, M. E., and Jannink, J. L. (2009). Genomic selection for crop improvement. *Crop Sci.* 49, 1–12. doi: 10.2135/cropsci2008.08.0512
- Ibba, M. I., Crossa, J., Montesinos-López, O. A., Montesinos-López, A., Juliana, P., Guzman, C., et al. (2020). Genome-based prediction of multiple wheat quality traits in multiple years. *Plant Genome* 13, 1–14. doi: 10.1002/tpg2.20034
- IWGSC (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191

- Jannink, J. L., Lorenz, A. J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genomics Proteomics* 9, 166–177. doi: 10.1093/bfgp/eq001
- Jarquín, D., Howard, R., Crossa, J., Beyene, Y., Gowda, M., Martini, J. W. R., et al. (2020). Genomic prediction enhanced sparse testing for multi-environment trials. *G3 Genes Genomes Genet.* 10, 2725–2739. doi: 10.1534/g3.120.401349
- Jia, Y., and Jannink, J. L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192, 1513–1522. doi: 10.1534/genetics.112.144246
- Jiang, J., Zhang, Q., Ma, L., Li, J., Wang, Z., and Liu, J. F. (2015). Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. *Heredity* 115, 29–36. doi: 10.1038/hdy.2015.9
- Juliana, P., Singh, R. P., Braun, H.-J., Huerta-Espino, J., Crespo-Herrera, L., Govindan, V., et al. (2020). Genomic selection for grain yield in the CIMMYT wheat breeding program—status and perspectives. *Front. Plant Sci.* 11:564183. doi: 10.3389/fpls.2020.564183
- Juliana, P., Singh, R. P., Singh, P. K., Crossa, J., Huerta-Espino, J., Lan, C., et al. (2017). Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. *Theor. Appl. Genet.* 130, 1415–1430. doi: 10.1007/s00122-017-2897-1
- Lado, B., Vázquez, D., Quincke, M., Silva, P., Aguilar, I., and Gutiérrez, L. (2018). Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theor. Appl. Genet.* 131, 2719–2731. doi: 10.1007/s00122-018-3186-3
- Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., Iwata, H., et al. (2011). *Advances in Agronomy*. Amsterdam: Elsevier.
- Maulana, F., Kim, K., Anderson, J. D., Sorrells, M. E., Butler, T. J., Liu, S., et al. (2021). Genomic selection of forage agronomic traits in winter wheat. *Crop Sci.* 61, 410–421. doi: 10.1002/csc2.20304
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). *Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps*. Available at: <https://www.genetics.org/content/157/4/1819.short> (Accessed January 12, 2021).
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Toledo, F. H., Pérez-Hernández, O., Eskridge, K. M., et al. (2016). A genomic bayesian multi-trait and multi-environment model. *G3 Genes Genomes Genet.* 6, 2725–2774. doi: 10.1534/g3.116.032359
- Montesinos-López, O. A., Montesinos-López, A., Luna-Vázquez, F. J., Toledo, F. H., Pérez-Rodríguez, P., Lillemo, M., et al. (2019). An R package for Bayesian analysis of multi-environment and multi-trait multi-environment data for genome-based prediction. *G3 Genes Genomes Genet.* 9, 1355–1369. doi: 10.1534/g3.119.400126
- Oury, F. X., Godin, C., Mailliar, A., Chassin, A., Gardet, O., Giraud, A., et al. (2012). A study of genetic progress due to selection reveals a negative effect of climate change on bread wheat yield in France. *Eur. J. Agron.* 40, 28–38. doi: 10.1016/j.eja.2012.02.007
- Pérez-Rodríguez, P., Gianola, D., González-Camacho, J. M., Crossa, J., Manès, Y., and Dreisigacker, S. (2012). Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3 Genes Genomes Genet.* 2, 1595–1605. doi: 10.1534/g3.112.003665
- Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. (2012b). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5, 103–113. doi: 10.3835/plantgenome.2012.06.0006
- Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J.-L. (2012a). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7:e32253. doi: 10.1371/journal.pone.0032253
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Available online at: <https://scholar.google.com/scholar?cluster=9441913529578809097&hl=en&andsdt=5,42&andsciodt=0,42> (accessed September 17, 2019).
- Randhawa, H. S., Asif, M., Pozniak, C., Clarke, J. M., Graf, R. J., Fox, S. L., et al. (2013). Application of molecular markers to wheat breeding in Canada. *Plant Breed.* 132, 458–471. doi: 10.1111/pbr.12057
- Rutkoski, J., Benson, J., Jia, Y., Brown-Guedira, G., Jannink, J.-L., and Sorrells, M. (2012). Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. *Plant Genome* 5, 51–61. doi: 10.3835/plantgenome.2012.02.0001
- Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Pérez, L. G., Crossa, J., et al. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 Genes Genomes Genet.* 6, 2799–2808. doi: 10.1534/g3.116.032888
- Rutkoski, J. E., Poland, J. A., Singh, R. P., Huerta-Espino, J., Bhavani, S., Barbier, H., et al. (2014). Genomic selection for quantitative adult plant stem rust resistance in wheat. *Plant Genome* 7, 1–10. doi: 10.3835/plantgenome.2014.02.0006
- Schulthess, A. W., Zhao, Y., Longin, C. F. H., and Reif, J. C. (2018). Advantages and limitations of multiple-trait genomic prediction for Fusarium head blight severity in hybrid wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 131, 685–701. doi: 10.1007/s00122-017-3029-7
- Sun, J., Rutkoski, J. E., Poland, J. A., Crossa, J., Jannink, J., and Sorrells, M. E. (2017). Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. *Plant Genome* 10, 1–12. doi: 10.3835/plantgenome.2016.11.0111
- Tadesse, W., Sanchez-Garcia, M., Gizaw Assefa, S., Amri, A., Bishaw, Z., Ogbonnaya, F. C., et al. (2019). Genetic gains in wheat breeding and its role in feeding the world. *Crop Breed. Genet. Genomics* 1. doi: 10.20900/cbgg2.0190005
- Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., et al. (2016). GAPIT version 2: an enhanced integrated tool for genomic association and prediction. *Plant Genome* 9, 1–9. doi: 10.3835/plantgenome.2015.11.0120
- Tester, M., and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822. doi: 10.1126/science.1183700
- VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., et al. (2009). Invited review: reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92, 16–24. doi: 10.3168/jds.2008-1514
- Wang, X., Li, L., Yang, Z., Zheng, X., Yu, S., Xu, C., et al. (2017). Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. *Heredity* 118, 302–310. doi: 10.1038/hdy.2016.87
- Wang, X., Xu, Y., Hu, Z., and Xu, C. (2018). Genomic selection methods for crop improvement: current status and prospects. *Crop J.* 6, 330–340. doi: 10.1016/j.cj.2018.03.001
- Ward, B. P., Brown-Guedira, G., Tyagi, P., Kolb, F. L., Van Sanford, D. A., Sneller, C. H., et al. (2019). Multi-environment and multitrait genomic selection models in unbalanced early-generation wheat yield trials. *Crop Sci.* 59, 491–507. doi: 10.2135/cropsci.2018.03.0189

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Gill, Halder, Zhang, Brar, Rai, Hall, Bernardo, Amand, Bai, Olson, Ali, Turnipseed and Sehgal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Multiple Wheat Genomes Reveal Novel *Gli-2* Sublocus Location and Variation of Celiac Disease Epitopes in Duplicated α -Gliadin Genes

Gwyneth Halstead-Nussloch^{1†}, Tsuyoshi Tanaka^{2,3†}, Dario Copetti^{1,4}, Timothy Paape^{1,5}, Fuminori Kobayashi², Masaomi Hatakeyama^{1,6}, Hiroyuki Kanamori², Jianzhong Wu², Martin Mascher^{7,8}, Kanako Kawaura⁹, Kentaro K. Shimizu^{1,9*} and Hirokazu Handa^{2,10*}

OPEN ACCESS

Edited by:

Peter Stephen Baenziger,
University of Nebraska System,
United States

Reviewed by:

Marinus J. M. Smulders,
Wageningen University and Research,
Netherlands
Heinrich Grausgruber,
University of Natural Resources
and Life Sciences Vienna, Austria

*Correspondence:

Hirokazu Handa
hirokazu@kpu.ac.jp
Kentaro K. Shimizu
kentaro.shimizu@uzh.ch

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 27 May 2021

Accepted: 12 August 2021

Published: 03 September 2021

Citation:

Halstead-Nussloch G, Tanaka T, Copetti D, Paape T, Kobayashi F, Hatakeyama M, Kanamori H, Wu J, Mascher M, Kawaura K, Shimizu KK and Handa H (2021) Multiple Wheat Genomes Reveal Novel *Gli-2* Sublocus Location and Variation of Celiac Disease Epitopes in Duplicated α -Gliadin Genes. *Front. Plant Sci.* 12:715985. doi: 10.3389/fpls.2021.715985

¹ Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, ² Institute of Crop Science, National Agriculture and Food Research Organization, Tsukuba, Japan, ³ Research Center for Advanced Analysis, National Agriculture and Food Research Organization, Tsukuba, Japan, ⁴ Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zürich, Zurich, Switzerland, ⁵ Brookhaven National Laboratory, Upton, NY, United States, ⁶ Functional Genomics Center Zurich, Zurich, Switzerland, ⁷ Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, ⁸ German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany, ⁹ Kihara Institute for Biological Research, Yokohama City University, Yokohama, Japan, ¹⁰ Graduate School of Life and Environmental Sciences, Kyoto Prefectural University, Kyoto, Japan

The seed protein α -gliadin is a major component of wheat flour and causes gluten-related diseases. However, due to the complexity of this multigene family with a genome structure composed of dozens of copies derived from tandem and genome duplications, little was known about the variation between accessions, and thus little effort has been made to explicitly target α -gliadin for bread wheat breeding. Here, we analyzed genomic variation in α -gliadins across 11 recently published chromosome-scale assemblies of hexaploid wheat, with validation using long-read data. We unexpectedly found that the *Gli-B2* locus is not a single contiguous locus but is composed of two subloci, suggesting the possibility of recombination between the two during breeding. We confirmed that the number of immunogenic epitopes among 11 accessions varied. The D subgenome of a European spelt line also contained epitopes, in agreement with its hybridization history. Evolutionary analysis identified amino acid sites under diversifying selection, suggesting their functional importance. The analysis opens the way for improved grain quality and safety through wheat breeding.

Keywords: α -gliadin, celiac disease epitopes, copy number variation, *Gli-2* loci, wheat (*Triticum aestivum* L.)

INTRODUCTION

Since its origin by allopolyploidization, bread wheat (*Triticum aestivum* L.) has become a staple crop, providing ~20% of the calories consumed globally (Shiferaw et al., 2013). Concentrated breeding efforts have increased yield such that the production of bread wheat reached 766 million tons in 2019 (FAOSTAT, 2021). Further selection has made wheat more palatable and increased the quality of desired end-use traits.

Wheat grains are typically processed into flour to make various breads and noodles. Much of the rheological quality of these products relies on gluten formation. Gluten is a complex of two protein families, glutenins and gliadins, which are storage proteins in wheat endosperm. Gliadins are classified into three groups, α -, γ -, and ω -gliadins, based on their electrophoretic mobility (Shewry and Halford, 2002). α -Gliadins are the most abundant gliadins and represent 15–30% of the wheat seed protein (Gu et al., 2004). Genes encoding α -gliadins are tandemly duplicated and form clusters within each *Gli-2* locus. They are located on the short arm of homoeologous chromosome group 6 in *Triticum* species (Qi et al., 2013; Ozuna et al., 2015). Because *T. aestivum* is allohexaploid (AABBDD), it contains three *Gli-2* loci called *Gli-A2*, *Gli-B2*, and *Gli-D2* (Payne, 1987). High allelic diversity, copy number variation, and expression differences in α -gliadins in bread wheat have been attributed to the combination of tandem and whole genome duplications (Salentijn et al., 2009; Noma et al., 2016; Huo et al., 2018). Although allelic diversity, gene copy number variation and other differences in α -gliadins may be linked to the phenotypic differences for the wheat flour qualities among the cultivars, little is known about precise genomic information for the *Gli-2* loci to provide a basis for comparison between cultivars. For example, the general function of α -gliadins in the breadmaking quality is well understood, but the role of individual α -gliadin genes is not entirely clear (Branlard et al., 2001; Brennan, 2009; Metakovsky et al., 2018). α -Gliadins are also the most common proteins that trigger an immune response in patients with celiac disease (CD), one of the widespread wheat-related health disorders (Scherf et al., 2016). The CD reaction is caused by the presence of a variety of peptide sequences called epitopes (Sollid et al., 2012; Juhász et al., 2018). The repetitive domain includes the DQ2.5-glia- α 1, DQ2.5-glia- α 2, and DQ2.5-glia- α 3 epitopes (Ozuna et al., 2015), and sometimes, these epitopes overlap to create a 33-mer peptide that is highly immunotoxic to celiac patients (Shan et al., 2002; Huo et al., 2018; Juhász et al., 2018). The immunotoxicity of the 33-mer region was verified by genome-editing (Sánchez-León et al., 2018). Although the three-dimensional structure of a short CD epitope bound to human HLA has been reported (Kim et al., 2004; Petersen et al., 2014, 2016), little is known about the higher-order structure of gliadin proteins because they aggregate in solutions (Urade et al., 2018). Sequence-based characterization of α -gliadin variation within modern hexaploid wheat cultivars will aid in breeding efforts to incorporate both desired end-use quality and lower reactivity for consumers.

Allopolyploidization and tandem duplication have made regions such as *Gli-2* difficult to characterize in terms of the genomic organization of and variation within multigene families found in bread wheat. Thus, most variation within α -gliadin gene sequences of different wheat accessions and related species has been detected using bacterial artificial chromosome (BAC) clones, transcriptome analysis, or low-coverage shotgun genome sequencing (Kawaura et al., 2012; Noma et al., 2016; Juhász et al., 2018). High resolution of the structure of homoeologous *Gli-2* loci has been described using long-read sequences, but in only one cultivar, Chinese Spring (CS; Huo et al., 2018).

Recently, advances in polyploid genomics enabled the high-quality genome assembly and polymorphism analysis of tandem duplications (Paape et al., 2016, 2018; Avni et al., 2017). Here, using chromosome-level assemblies for 11 accessions including elite bread wheat cultivars and a spelt wheat line in the framework of the “10+ Wheat Genomes Project” (Walkowiak et al., 2020), we began to address the question of global variation in both the structure of and polymorphism within *Gli-2* loci among multiple cultivars.

MATERIALS AND METHODS

Sequence Resources

Reference-quality genome assemblies for 9 bread wheat accessions, ArinaLrFor, CDC Landmark, CDC Stanley, Jagger, Julius, LongReach Lancer, Mace, Norin 61, SY Mattis and one spelt accession, PI190962, released by the “10+ Wheat Genomes Project” (Walkowiak et al., 2020), were accessed through IPK, Germany.¹ We also used the RefSeq v1.0 assembly of CS (International Wheat Genome Sequencing Consortium (IWGSC) et al., 2018), which is available at INRAE, France.²

Identification of *Gli-2* Loci and α -Gliadin Sequences

To identify the location of the *Gli-2* loci, BLAST searches were conducted against chromosome assemblies for homoeologous group 6 of the eleven accessions using the α -gliadin gene sequences AS2 and AS7 (for chromosome 6A); AS3, AS4, AS5 and AS6 (for 6B); and AS1, AS8, AS9, AS10, and AS11 (for 6D) as queries (Noma et al., 2016). From the BLAST results, regions with an *e*-value = 0 and composed of a single exon were selected as candidates for α -gliadin gene copies. The regions were translated into amino acid sequences. Sequences not starting with a methionine residue were discarded as incomplete gene fragments. Sequences that were too diverged based on the sequence alignment or phylogenetic tree were also omitted. Finally, we constructed a codon-based alignment of gliadin gene copies using MUSCLE in MEGA (Kumar et al., 2018). Hi-C data and the alignments of CDC Landmark Oxford Nanopore Technologies (ONT) long-read data were obtained from Walkowiak et al. (2020). The gene coverage values of the long-read alignments were obtained with SAMtools v1.0 (Li et al., 2009) and BEDtools v2.29.0 (Quinlan and Hall, 2010). Read alignments were visualized with IGV v2.8.2 (Robinson et al., 2017).

The evolutionary history of the gene family was inferred from the 429 α -gliadin sequences identified above. Codon positions included were 1st + 2nd + 3rd + Non-coding. All positions containing gaps and missing data were eliminated (complete deletion option). The final dataset contained a total of 524 positions. The tree was estimated using the neighbor-joining method (Saitou and Nei, 1987), and evolutionary distances were computed using the Kimura 2-parameter method (Kimura, 1980).

¹<https://wheat.ipk-gatersleben.de/>

²<https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies>

and expressed as the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 2.25). Support for the tree topology was estimated using the bootstrap test with 1,000 replicates and was calculated as the percentage of replicate trees in which the associated taxa clustered together (Felsenstein, 1985). The tree was drawn to scale, with the units for branch lengths being the same as those of the evolutionary distances used to infer the phylogenetic tree.

Celiac Disease Epitope Search and Site Selection Analysis

Using the amino acid sequences of α -gliadin copies without the last stop codon, we searched all sequences for the presence of nine canonical amino acid epitopes previously shown to induce an immunogenic reaction (Sollid et al., 2012; Ozuna et al., 2015).

To test for amino acid sites likely to be under positive selection in the α -gliadin gene family, only full-length sequences were considered for a conservative analysis. Gaps present at the same position in all three *Gli-2* loci and sequences containing premature stop codons were discarded. Sequences were also removed if they had no terminal stop codon or were not composed of multiples of three nucleotides, implying frameshifts. Last, sites in regions that were difficult to align (the polyglutamine regions) were not considered, as the uncertain alignments may produce false positive signals. For the selection analysis, a phylogenetic method was applied. First, the most likely phylogenetic tree was estimated using nucleotide alignment and a general time reversible (GTR) + invariant + gamma model in MrBayes (Ronquist et al., 2012). Then, the likelihood of that tree was calculated under different codon substitution models by estimating the non-synonymous and synonymous substitution rate ratios ($\omega = d_N/d_S$) for each codon within the alignment. The value of ω indicates the type of selection: $\omega < 1$ indicates negative selection, $\omega = 0$ indicates neutral evolution, and $\omega > 1$ indicates positive selection. A likelihood ratio test (LRT) was run between two nested codon substitution models, a null and an alternative model, to determine whether the alternative model of positive selection was supported. The null model (M7) did not allow for sites under positive selection while the alternative model (M8) did allow for positive selection (Yang et al., 2000). Last, the posterior probability of a specific site being under positive selection was estimated using Bayesian empirical Bayes (BEB) (Yang et al., 2005). Sites with a probability > 95% were considered significant. The likelihoods of the codon substitution models and posterior probability calculations were implemented in the CODEML program of the software package PAML4 (Yang, 2007).

RESULTS

Location and Validation of *Gli-2* Loci in Accessions

We identified α -gliadin gene copies within 11 wheat assemblies: the 10 reference-quality pseudomolecule assemblies (Walkowiak et al., 2020) plus CS RefSeq v1.0 (International Wheat

Genome Sequencing Consortium (IWGSC) et al., 2018). We first examined the chromosomal positions of the α -gliadin copies. Copies that mapped to chromosome 6A in the 11 wheat accessions were assigned as *Gli-A2* and were located in single region on the short arm, as expected (Table 1 and Supplementary Table 1). The only exception was *Gli-A2* of CDC Landmark, which was split into 2 subloci 7 Mb apart from each other. Similarly, sequences in *Gli-D2* mapped to the expected region on chromosome 6D in nine reference-quality assemblies. In LongReach Lancer and CS, we could not identify α -gliadin copies on chromosome 6D; however, those found in scaffolds that were not anchored to a chromosome (chrUn) were assigned to *Gli-D2* following the suggestion of Juhász et al. (2018). Surprisingly, the copies found on chromosome 6B showed that the *Gli-B2* locus was clearly split into 2 subloci in all accessions. We called them *Gli-B2-1* and *Gli-B2-2*, and they were 12–21 Mb apart from each other on chromosome 6B (Table 1 and Supplementary Table 1). The uniformity of the Hi-C signal along the whole *Gli-B2* region and its flanking regions further supported that the bipartite structure of *Gli-B2* was not an assembly artifact (Supplementary Figure 1) (Shimizu et al., 2020; Walkowiak et al., 2020). The position of this second locus relative to the well-described locus at ~43 Mb on chromosome 6B in CS has not been described before, although previous studies mention two sequences that mapped outside that region (Huo et al., 2018; Juhász et al., 2018). The consistency in the Hi-C maps observed among all assemblies supports that *Gli-B2* is composed of two parts and opens the possibility of exploiting genetic recombination for breeding purposes.

Although some assemblies showed the subdivision or translocation of several α -gliadin genes compared to those of other accessions, we interpreted them with caution. In Jagger, α -gliadin sequences mapped to a third region (*Gli-B2-3*) located at the end of the long arm of chromosome 6B. We also found that the sublocus *Gli-B2-1* of LongReach Lancer and CDC Landmark was further split into two parts. However, the Hi-C signal of intra- and interchromosomal interactions for these accessions suggested potential misassembly within these regions (Supplementary Figures 1, 2). We note that the 20 Mb regions flanking *Gli-B2-1* in CDC Landmark were highly concordant with those in CS, but they were not concordant with those in LongReach Lancer (Supplementary Figure 3). Because the assembly structure and orientation of CS was also supported by additional evidence (International Wheat Genome Sequencing Consortium (IWGSC) et al., 2018), this suggested that the rearrangement in CDC Landmark may represent true biological variation. The location and orientation of these subloci remain interesting cases for further validation to distinguish biological rearrangement from assembly errors.

Next, we checked the accuracy of the assemblies around each single α -gliadin gene copy. We utilized the long-read sequence data from ONT for CDC Landmark that was previously used to validate the assembly (Walkowiak et al., 2020). Though often flanked by assembly gaps, the sequence at and immediately adjacent to each α -gliadin gene copy was supported by continuous alignments of several long reads (Supplementary Figure 4), implying a gene-level correctness

TABLE 1 | Genomic positions of *Gli-2* loci in 11 wheat accessions.

Accession	<i>Gli-A2</i>		<i>Gli-B2</i>				<i>Gli-D2</i>			
	Positions*	Copy no.**	<i>Gli-B2-1*</i>	Copy no.**	<i>Gli-B2-2*</i>	Copy no.**	<i>Gli-B2-3*</i>	Copy no.**	Positions*	Copy no.**
ArinaLrFor	25.5–27.7	30 (14)	44.5–45.1	6 (4)	63.0	2 (1)			28.1–28.5	9 (5)
CDC Landmark	25.8–26.4 and 33.0–33.1	11 (9)	43.4–43.6 and 49.5–49.7	12 (6)	63.2	2 (1)			27.1–27.5	9 (5)
CDC Stanley	26.1–27.0	11 (9)	45.3–46.0	12 (7)	65.2	2 (1)			28.2–28.7	10 (6)
Jagger	24.7–25.6	11 (9)	43.6–44.4	9 (3)	65.0	2 (1)	654.9–655.2	6 (4)	26.7–26.8	3 (1)
Julius	24.7–25.5	10 (8)	46.6–47.1	6 (3)	64.7	2 (0)			27.0–27.4	9 (5)
LongReach Lancer	24.6–25.4	11 (9)	35.5–35.7 and 45.7–46.0	14 (6)	58.2	2 (1)			chrUn	11 (8)
Mace	24.8–25.6	11 (9)	43.5–44.3	13 (7)	63.8	2 (1)			26.8–27.3	11 (7)
Norin 61	25.8–28.2	33 (15)	42.3–43.0	14 (8)	61.2	2 (1)			26.7–27.2	11 (6)
SY Mattis	24.0–25.9	23 (8)	44.6–45.2	6 (2)	61.8	2 (0)			26.6–27.1	9 (5)
PI190962 (spelt)	25.8–28.2	25 (10)	41.3–41.7	5 (2)	60.4	2 (1)			26.4–26.9	11 (6)
Chinese Spring	24.9–25.6	9 (7)	43.4–44.1	8 (3)	62.7	2 (1)			chrUn	18 (12)

Genomic positions (in Mb) and copy number of α -gliadin genes that mapped within chromosome 6A, 6B, or 6D of each accession were assigned to *Gli-A2*, *Gli-B2*, and *Gli-D2*, respectively. Within LongReach Lancer and Chinese Spring, copies that mapped to chrUn were assigned to *Gli-D2*. *Genomic positions (Mb) in the pseudomolecules of the corresponding chromosomes. **The number of intact genes encoded in this region is shown in parentheses.

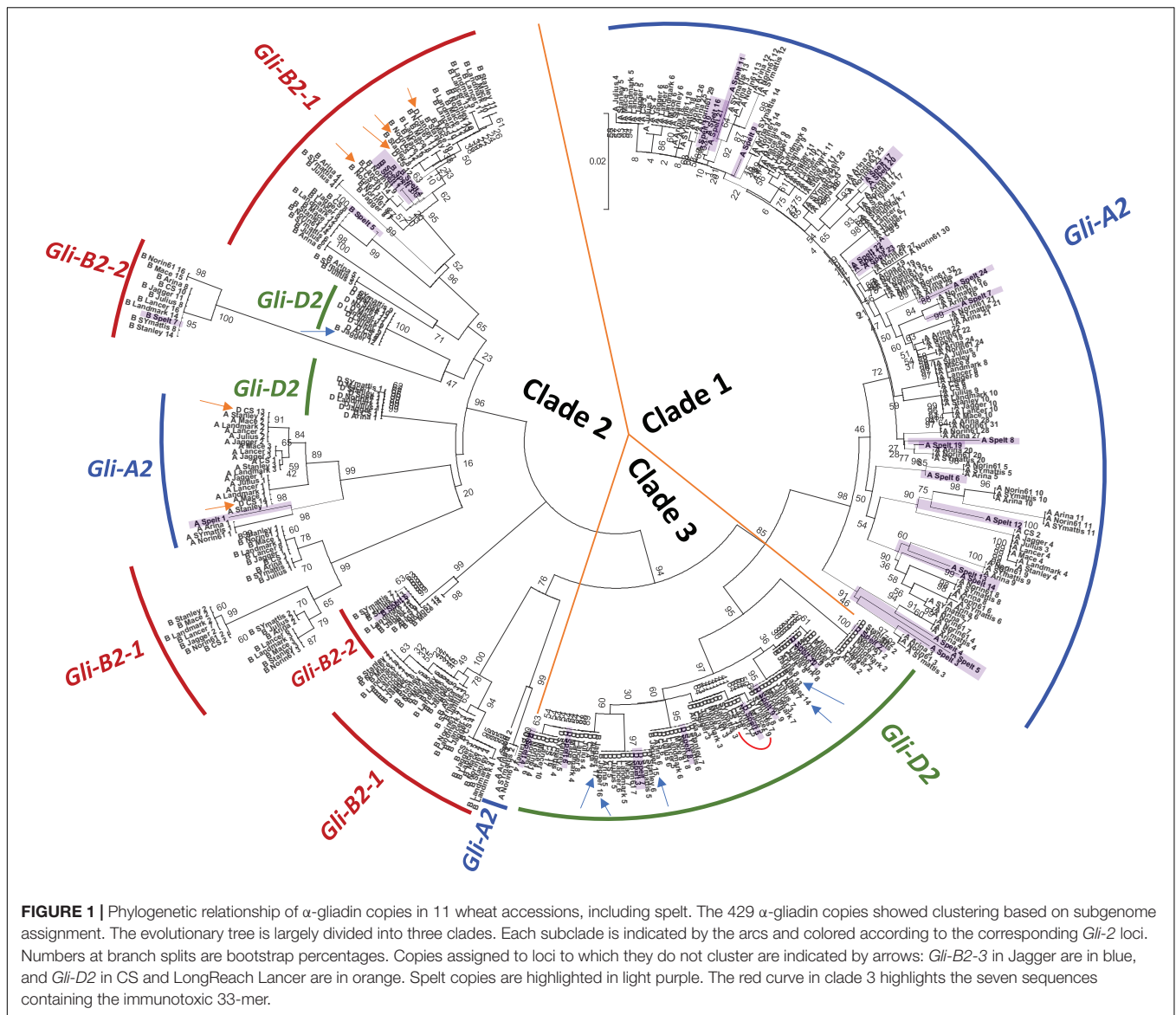
of each model. The different coverage seen among copies, including those in close proximity, hinted at potential collapses of paralogous copies into a single gene (**Supplementary Figure 4B**) or the separate assembly of allelic heterozygous copies. To address this possibility, we compared the coverage of the ONT alignments for each α -gliadin gene in the assembly to the median genome-wide gene coverage (32.23 genome equivalents). Of the 34 copies that we manually annotated in CDC Landmark, seven (20.5%) had a mean coverage that clearly deviated from that of other copies. As a comparison, the coverage of the three *ADH* homoeologs (chosen as a single-copy gene reference) was well within the genome-wide value (**Supplementary Table 1**). While the three α -gliadin copies at high coverage likely represent collapsed paralogs, the four genes at lower coverage may be haplotype-specific assemblies of heterozygous allelic copies. The long-read data suggested that the assembled α -gliadin sequences were correctly identified, although the exact copy number of ~20% of them may be different.

The number of assembled α -gliadin genes within each *Gli-2* locus is reported in **Table 1**. While most accessions possessed approximately 11 α -gliadin copies in *Gli-A2*, the accessions ArinaLrFor, Norin 61, SY Mattis, and PI190962 had two to three times as many copies (**Table 1**). We identified 13–17 copies in *Gli-B2* in most accessions, while ArinaLrFor, Julius, SY Mattis and the European spelt PI190962 had only half the number of copies compared to the other accessions (**Table 1**). For *Gli-D2*, there were approximately 10 copies in most accessions (**Table 1**). An extremely high or low copy number for *Gli-B2* and *Gli-D2* in Jagger, respectively, was possibly an assembly error, as described above. Subsequent analyses in this paper will use the assignment to a particular *Gli-2* locus based on previously published assemblies for consistency.

Phylogenetic Analysis of α -Gliadin Copies

We then assessed the relationship between all α -gliadin copies identified in the 11 accessions using phylogenetic analysis. According to the clustering pattern, α -gliadin copies were classified into three main clades named 1, 2, and 3 (**Figure 1**). Clades 1 and 3 showed a compact structure and included copies mostly from *Gli-A2* to *Gli-D2*, respectively. In *Gli-D2*, unlike other subgenome loci, there was little difference in copy number between accessions and the genetic distances between branches were shorter. Limited allelic diversity at the *Gli-D2* locus is consistent with the lower diversity of the coding sequences in the D subgenome (International Wheat Genome Sequencing Consortium (IWGSC), 2014; Jordan et al., 2015; Walkowiak et al., 2020). Clade 2 mostly contained copies from *Gli-A2* and *Gli-D2*, although with weak branch support (**Figure 1**). As mentioned above, we found 2 subloci in *Gli-B2*, i.e., *Gli-B2-1* and *Gli-B2-2* (**Table 1**). In the phylogenetic tree (**Figure 1**), the two gliadin sequences encoded in *Gli-B2-2* formed subclades distinct from other sequences in *Gli-B2-1*, indicating that the split of *Gli-B2* was shared among all wheat accessions and that the genes in the two subloci experienced different histories. These data further support the bipartite structure of *Gli-B2*.

The α -gliadin copies that mapped to Jagger *Gli-B2-3* clustered in clade 3, which is composed of *Gli-D2* copies (**Figure 1**, blue arrows). This is consistent with the possible misassembly from *Gli-D2* to the end of chromosome 6BL in this cultivar (see also the previous section; **Supplementary Figure 2**). If we reassign these sequences as a part of *Gli-D2*, the copy numbers of *Gli-B2* and *Gli-D2* in Jagger are closer to the average copy number found within the other accessions. In the



case of LongReach Lancer and CS, we assigned all α -gliadin copies in chrUn as copies of *Gli-D2* following the suggestion of Juhász et al. (2018). However, we found that several copies clustered with those assigned to *Gli-A2* or *Gli-B2* in the other accessions (Figure 1, orange arrows). Despite these potential misclassifications, we were able to show that there were clear variations among accessions.

Among the accessions with the largest differences in copy numbers, we observed distinct clustering patterns. Branches from accessions with the highest copy number for *Gli-A2*, such as ArinaLrFor, Norin 61, SY Mattis and PI190962, were clearly separated from the branches of the other seven accessions. Similarly, we found distinct clusters containing copies of *Gli-B2-1* from ArinaLrFor, Julius, and SY Mattis. These three accessions, in addition to PI190962, contained the lowest copy number within this locus. These examples highlight potential differences in evolutionary and/or breeding history between

accessions and that the gene duplications or losses in some cultivars did not originate independently but were likely from a common ancestor.

Focusing on the spelt wheat, PI190962, we observed no clear association with the other accessions for *Gli-A2* and *Gli-B2-1*. Rather, most of the copies in PI190962 formed their own branches or small clusters. This was not the case for *Gli-D2*, where the PI190962 copies were positioned on the same branches as those for other bread wheats. Interestingly, the α -gliadin copies in *Gli-B2-2* in PI190962 also clustered with those of the other accessions (Figure 1, purple highlight). The spelt accession, PI190962, used in this study is a Central European spelt, which has been suggested to have originated from the introgression of a hulled tetraploid emmer wheat into bread wheat during the migration of bread wheat from the Fertile Crescent to Europe. Therefore, the A and B subgenomes between bread wheat and European spelt had higher sequence divergence,

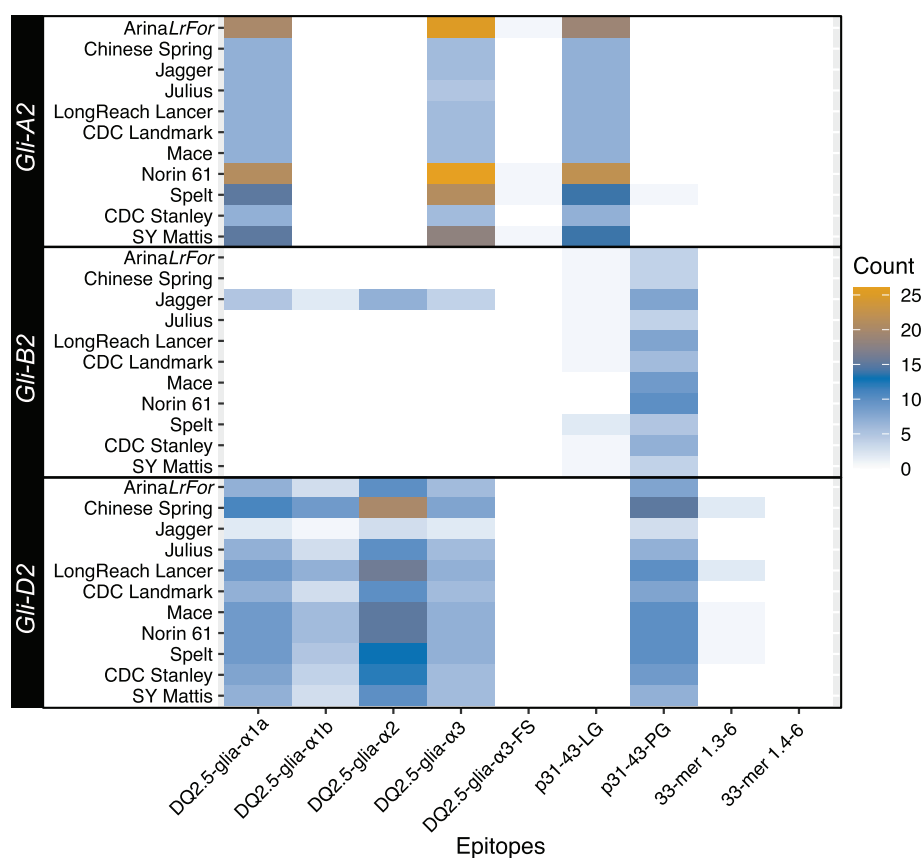


FIGURE 2 | Celiac disease (CD) epitope quantification within α -gliadin copies. The frequency of canonical CD epitopes varies between accessions and homoeologous chromosomes.

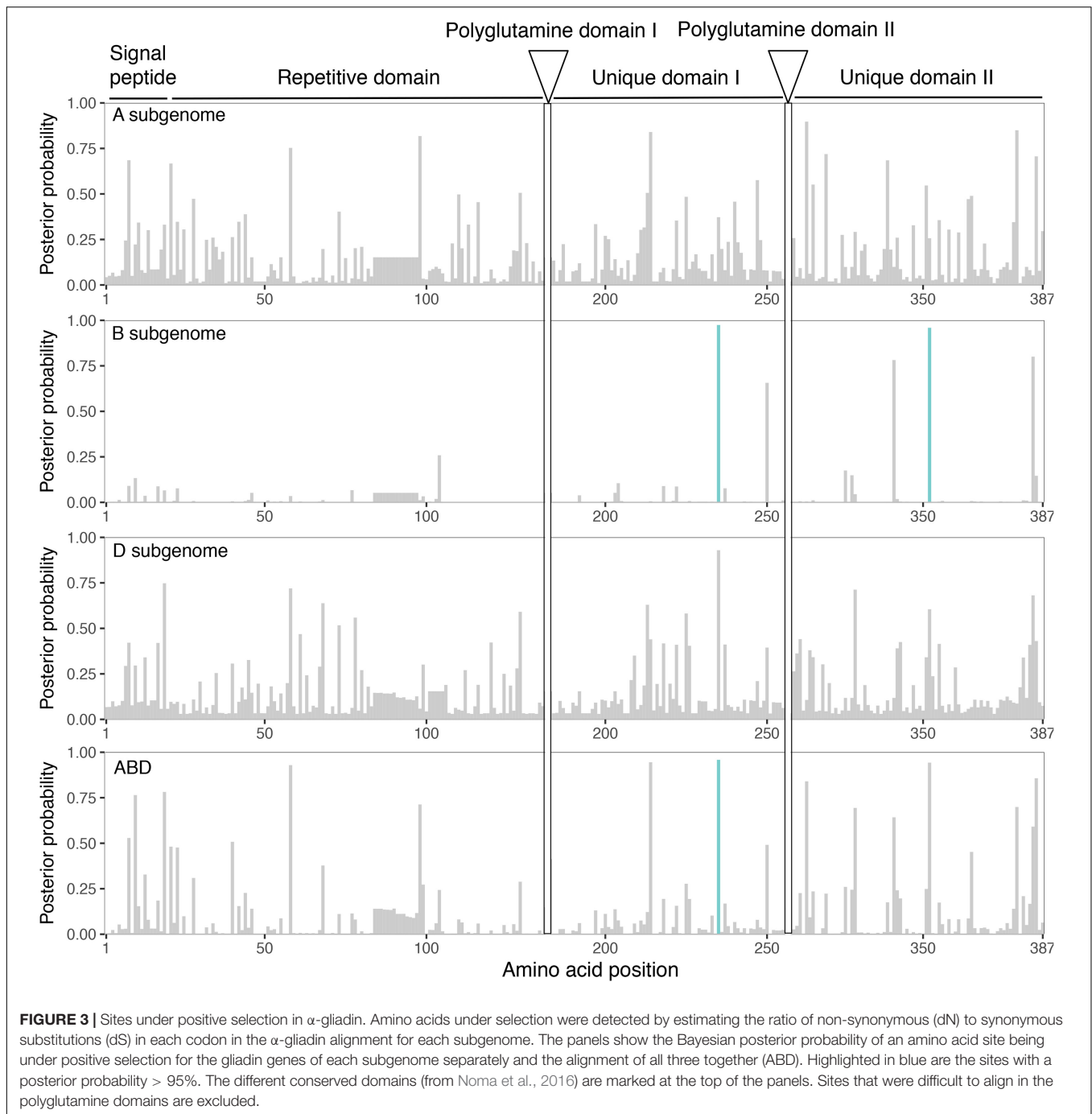
while the D subgenome showed greater sequence similarity (Blatter et al., 2004; Dvorak et al., 2012). Our observation of the separation of α -gliadin sequences in PI190962 from those of other bread wheats in *Gli-A2* and *Gli-B2-1* supports that this Central European spelt accession had an introgressed origin with a tetraploid emmer wheat, which was recently shown to be distinct from the origins of Iberian spelt (Abrouk et al., 2021). This result also indicates that the introgressed loci from chromosome 6B of emmer wheat may be confined to the region encoding *Gli-B2-1*, further supporting a different evolutionary history for the two *Gli-B2* loci identified in this study.

Celiac Disease Epitope Copy Number and Positive Selection in α -Gliadins

Specific epitopes found in α -gliadins can induce reactions in patients with CD and gluten intolerance. Therefore, the search for new alleles and/or copy number variations that may cause weaker or no reaction is beneficial in breeding programs. Among the amino acid sequences produced by α -gliadin genes from the 11 wheat accessions, we found polymorphic sites within three major immunogenic regions, p31-43, the 33-mer, and the DQ2.5-glia- α 3 peptide, using the established nomenclature (Sollid et al., 2012; Ozuna et al., 2015). The presence of epitope sequences

showed a subgenome-specific pattern within the 11 accessions (Figure 2), and the count of CD epitopes in each accession mirrored the total α -gliadin copy number present in each locus (Table 1). *Gli-A2* contained mostly DQ2.5-glia- α 1b, DQ2.5-glia- α 3 and p31-43-LG epitopes. Variants of the latter two epitopes were also present, but at low frequency, in four accessions. The B subgenome encoded the fewest epitopes, the highest proportion of which were p31-43-PG. Among all accessions, the largest variety of CD epitopes was present in *Gli-D2* and included several that overlapped in a single gene copy. The toxic 33-mer sequence that contains six epitopes (33-mer 1.3-6) was found in the *Gli-D2* sequence of 5 accessions, including once in PI190962 and twice each in CS and LongReach Lancer (Figure 1, red curve and Figure 2).

In general, a single known epitope sequence was not found in the genes of all three subgenomes and sequences with multiple overlapping epitopes were restricted to *Gli-D2*. For example, DQ2.5glia- α 1a was present in both *Gli-A2* and *Gli-D2*. The sequence encoding both DQ2.5glia- α 1a and DQ2.5glia- α 2 (PFPQPQLPYPQ) was found only in *Gli-D2* due to a P to S substitution (PFPQPQLPYSQ) in *Gli-A2*. No DQ2.5-glia- α -type epitopes were present in *Gli-B2*, except for the potential misassembly or translocation of regions from chromosomes 6D to 6B in Jagger. The patterns we observed reflected those of



previous studies reporting the presence of specific epitopes in the subgenomes of hexaploid wheat (van Herpen et al., 2006; Salentijn et al., 2009; Sollid et al., 2012; Ozuna et al., 2015; Noma et al., 2016; Juhász et al., 2018).

Global prevalence of CD has increased (Singh et al., 2018) and this has been attributed, by some, to modern breeding practices. Due to its hybridization history, spelt wheats contain different gliadin and glutenin contents and has been subject to less intensive selection than modern bread wheats (Dubois et al., 2016; Escarnot et al., 2018), prompting the idea it could

be less reactive for consumers. We observed that the numbers and distribution patterns of the immunogenic epitopes in the particular spelt accession, PI190962, were similar to those of other bread wheat accessions (Figure 2), including one copy of the 33-mer peptide that was identified in *Gli-D2*. Although the study of Asian and other spelts (Blatter et al., 2004; Dvorak et al., 2012) would be necessary to draw conclusions about spelt diversity, the data from this single accession of spelt did not support the claim that spelt (as a species) could produce weaker reactions in people with CD, in agreement with previous genetic

studies (Ozuna et al., 2015; Dubois et al., 2016; Ruiz-Carnicer et al., 2019). Recent studies investigating overall protein and gluten content of both modern and old hexaploid wheat as well as “ancient” varieties including spelt, emmer, and einkorn showed no conclusive role of modern breeding techniques in the increased prevalence of CD. Rather, they exemplified the high variability of gluten content between all varieties, new and old, and reiterate the importance of environmental factors in overall protein content of wheat and its relatives (Escarnot et al., 2018; Geisslitz et al., 2019; Call et al., 2020; Pronin et al., 2021). Our results on the genetic variability are in line with these protein-based studies and, taken together, show the tools to identify low immunoreactive varieties are well developed. These studies not only proposed suitable varieties for further breeding already but also motivate a more comprehensive characterization of wheat and its relatives to tap into existing variability for breeding (Shewry, 2018).

We used a method to identify selection on amino acid-changing substitutions (PAML; Yang, 2007). This method estimates the ratio of amino-acid replacement mutations (non-synonymous substitutions, dN) compared with synonymous substitutions (dS). When the dN:dS ratio is greater than 1, it indicates positive or diversifying selection. Many positions showed a posterior probability higher than 0.75. Among them, in *Gli-B2*, we found two codon positions that were above the 95% significance level: one in unique domain I and another in unique domain II (Figure 3). When all *Gli-2* loci were analyzed together, the position in unique domain I remained significant (Figure 3 and Supplementary Table 2). The other amino acid position that was significant in *Gli-B2* domain II was just below the threshold when A, B and D were analyzed together (Figure 3).

DISCUSSION

The importance of bread wheat in human nutrition and its role in disease warrant the characterization of genetic and structural variation within the gene family encoding gliadin, which forms the gluten protein structure together with glutenin. However, this research has been challenging due to the complexity of the loci caused by tandem and homoeologous duplications. Here, we characterized the diversity of α -gliadin gene copies and their organization within *Gli-2* loci in chromosome-scale assemblies of 11 globally distributed bread and spelt wheat accessions. Long-read data supported that the assembled gliadin coding regions were correct, and 80% of them were assembled as a single copy with high confidence. The remaining 20% may possibly be collapsed, highly similar paralogs or independently assembled alleles of a gene copy. Unexpectedly, we found a bipartite structure of the *Gli-B2* loci in all assemblies, which was supported by Hi-C data and evolutionarily supported by phylogenetic analysis. This suggests that further expansion of the variation at the gliadin locus through chromosomal recombination using the segregation of these subloci may be applicable for future wheat breeding. Using the PAML method, we detected amino acid positions that were under diversifying selection, suggesting that polymorphisms at these positions may be relevant for functional

differences, such as those involved in interactions with glutenins (Li et al., 2014). This warrants further functional validation via amino acid substitution experiments.

Previous reports describe the subgenome specificity of sequences with CD epitopes, and those that cause the strongest immune cell reactions occur mostly in the A and D subgenomes and their respective progenitors. On the other hand, the wheat B subgenome, barley and several other Triticeae species contain epitopes that produce relatively weak responses from their α -gliadin and related proteins (van Herpen et al., 2006; Juhász et al., 2018). Our results not only reflect this subgenome specificity but also show that epitopes causing gluten-related reactions are unevenly distributed among accessions covering a wide range of wheat diversity (Walkowiak et al., 2020). The D subgenome is the only identified source of the toxic 33-mer epitope within bread wheat, and its presence has been detected at low frequency in the germplasm of the D progenitor *Aegilops tauschii* (Schaart et al., 2021). Current efforts to incorporate this knowledge into breeding safer varieties include the generation of synthetics and *Gli-D2* deletion lines (Camerlengo et al., 2017; Li et al., 2018), the development of probes to quickly confirm the presence of reactive epitopes (Dubois et al., 2017), and the genome-editing to reduce the immunotoxic 33-mer (Sánchez-León et al., 2018). Our study can inform these efforts. Our results show the reduced frequency of reactive epitopes in some accessions but also show that reactive epitopes are present in spelt, which is consistent with a previous study (Escarnot et al., 2018), indicating that detailed cultivar-specific analysis is needed. While the immunogenic effects of many of the polymorphic epitopes have not been directly tested, our main findings indicate that resources for breeding less reactive wheat are already present in elite germplasm.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: IPK, Germany (<https://wheat.ipk-gatersleben.de/>) INRAE, France (<https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies>).

AUTHOR CONTRIBUTIONS

KKS and HH conceived the study. TT, FK, HK, JW, and KK performed the identifications and phylogenetic analyses of α -gliadin genes. GH-N, DC, TP, and MH performed the epitope identification and evolutionary analyses within α -gliadins. DC and MM performed and curated the bioinformatic data. KK, KKS, and HH acquired the funding. GH-N, DC, KKS, and HH wrote the original draft. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by JST CREST (JPMJCR16O3), the Swiss National Science Foundation 31003A_182318, MEXT

KAKENHI The Birth of New Plant Species (JP16H06469), the University of Zurich Research Priority Program Evolution in Action to KKS, the National Agriculture and Food Research Organization (NARO) Vice President Fund to HH, and the NBRP Genome Information Upgrading Project 2017, AMED to KK and KKS.

ACKNOWLEDGMENTS

We thank Sean Walkowiak and Curtis Pozniak (Crop Development Centre, University of Saskatchewan, Saskatoon, Saskatchewan, Canada) for sharing the long-read alignments, and the IT Support Group D-HEST, ETH Zürich, for computational support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.715985/full#supplementary-material>

Supplementary Figure 1 | Large-scale chromosome organization around the *Gli-B2* loci in 11 accessions. Hi-C contact matrices of chromosome 6B at 0–100 Mb show a strong signal on the diagonal and a relatively even gradient

perpendicular to the diagonal, indicating that the large-scale structure of assemblies is correct. Dashed lines represent the position of each sub locus identified in *Gli-B2*.

Supplementary Figure 2 | Interchromosomal interactions between chromosomes 6B and 6D for each accession. Interactions between these chromosomes are weak apart from small regions that show dark vertical lines (e.g., in Jagger, Landmark, Mace), representing evidence of possible misassembly.

Supplementary Figure 3 | Alignment of LongReach Lancer and CDC Landmark chromosome 6B to that of Chinese Spring (CS). In the region from 0 to 100 Mb on 6B, LongReach Lancer may have inversions compared with CS. On the other hand, though interrupted by assembly gaps and indels, the region is assembled in CDC Landmark with the same orientation as that in CS.

Supplementary Figure 4 | Validation of the *Gli-2* loci with long-read data. Oxford Nanopore Technologies long reads of CDC Landmark were aligned to the short read-based assembly of the same variety (Walkowiak et al., 2020). Alignments at the *Gli-2* loci were inspected for overall structure [(A), *Gli-A2a* on chr6A 25.7–26.5 Mb] and at the single-gene scale [(B) *Gli-B2-1a* on chr6B]. Despite the presence of assembly gaps, the sequence of the coding regions was well supported by multiple reads.

Supplementary Table 1 | Gene copy name, chromosome with start and stop position, gene annotation (if available) from Walkowiak et al. (2020), and number of amino acids (if full length) information for all copies identified within the 11 accessions.

Supplementary Table 2 | Posterior probabilities and estimated omega (ω) values for all amino acid sites in gliadin alignments for each subgenome analyzed separately and together.

REFERENCES

- Abrouk, M., Athiyannan, N., Müller, T., Pailles, Y., Stritt, C., Roulin, A. C., et al. (2021). Population genomics and haplotype analysis in spelt and bread wheat identifies a gene regulating glume color. *Commun. Biol.* 4, 1–11. doi: 10.1038/s42003-021-01908-6
- Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S. O., Gundlach, H., et al. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* 357, 93–97. doi: 10.1126/science.aa.n0032
- Blatter, R. H. E., Jacomet, S., and Schlumberg, A. (2004). About the origin of European spelt (*Triticum spelta* L.): allelic differentiation of the HMW Glutenin B1-1 and A1-2 subunit genes. *Theor. Appl. Genet.* 108, 360–367. doi: 10.1007/s00122-003-1441-7
- Branlard, G., Dardevet, M., Saccomano, R., Lagoutte, F., and Gourdon, J. (2001). Genetic diversity of wheat storage proteins and bread wheat quality. *Euphytica* 119, 59–67. doi: 10.1023/A:1017586220359
- Brennan, C. (2009). Gliadin and Glutenin: the unique balance of wheat quality. *Int. J. Food Sci. Technol.* 44, 1865–1866. doi: 10.1111/j.1365-2621.2007.01703.x
- Call, L., Kapeller, M., Grausgruber, H., Reiter, E., Schoenlechner, R., and D'Amico, S. (2020). Effects of species and breeding on wheat protein composition. *J. Cereal Sci.* 93:102974. doi: 10.1016/j.jcs.2020.102974
- Camerlengo, F., Sestili, F., Silvestri, M., Colaprico, G., Margiotta, B., Ruggeri, R., et al. (2017). Production and molecular characterization of bread wheat lines with reduced amount of α -type gliadins. *BMC Plant Biol.* 17:248. doi: 10.1186/s12870-017-1211-3
- Dubois, B., Bertin, P., and Mingeot, D. (2016). Molecular diversity of α -gliadin expressed genes in genetically contrasted spelt (*Triticum aestivum* ssp. *spelta*) accessions and comparison with bread wheat (*T. aestivum* ssp. *aestivum*) and related diploid *Triticum* and *Aegilops* species. *Mol. Breeding* 36:152. doi: 10.1007/s11032-016-0569-5
- Dubois, B., Bertin, P., Muhovski, Y., Escarnot, E., and Mingeot, D. (2017). Development of TaqMan probes targeting the four major celiac disease epitopes found in α -gliadin sequences of spelt (*Triticum aestivum* ssp. *spelta*) and bread wheat (*Triticum aestivum* ssp. *aestivum*). *Plant Methods* 13:72. doi: 10.1186/s13007-017-0222-2
- Dvorak, J., Deal, K. R., Luo, M.-C., You, F. M., von Borstel, K., and Dehghani, H. (2012). The origin of spelt and free-threshing hexaploid wheat. *J. Hered.* 103, 426–441. doi: 10.1093/jhered/esr152
- Escarnot, E., Gofflot, S., Sinnaeve, G., Dubois, B., Bertin, P., and Mingeot, D. (2018). Reactivity of gluten proteins from spelt and bread wheat accessions towards A1 and G12 antibodies in the framework of celiac disease. *Food Chem.* 268, 522–532. doi: 10.1016/j.foodchem.2018.06.094
- FAOSTAT (2021). Available online at: <http://www.fao.org/faostat/en/#data/QC> (accessed February 8, 2021).
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678
- Geisslitz, S., Longin, C. F. H., Scherf, K. A., and Koehler, P. (2019). Comparative study on gluten protein composition of ancient (einkorn, emmer and spelt) and modern wheat species (Durum and Common Wheat). *Foods* 8:409. doi: 10.3390/foods8090409
- Gu, Y. Q., Crossman, C., Kong, X., Luo, M., You, F. M., Coleman-Derr, D., et al. (2004). Genomic organization of the complex α -gliadin gene loci in wheat. *Theor. Appl. Genet.* 109, 648–657. doi: 10.1007/s00122-004-1672-2
- Huo, N., Zhu, T., Altenbach, S., Dong, L., Wang, Y., Mohr, T., et al. (2018). Dynamic evolution of α -Gliadin prolamin gene family in homeologous genomes of hexaploid wheat. *Sci. Rep.* 8:5181. doi: 10.1038/s41598-018-23570-5
- International Wheat Genome Sequencing Consortium (IWGSC) (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345:1251788. doi: 10.1126/science.1251788
- International Wheat Genome Sequencing Consortium (IWGSC), Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191
- Jordan, K. W., Wang, S., Lun, Y., Gardiner, L.-J., MacLachlan, R., Hucl, P., et al. (2015). A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homeologous genomes. *Genome Biol.* 16:48. doi: 10.1186/s13059-015-0606-4

- Juhász, A., Belova, T., Florides, C. G., Maulis, C., Fischer, I., Gell, G., et al. (2018). Genome mapping of seed-borne allergens and immunoresponsive proteins in wheat. *Sci. Adv.* 4:eaar8602. doi: 10.1126/sciadv.aar8602
- Kawaura, K., Wu, J., Matsumoto, T., Kanamori, H., Katagiri, S., and Ogihara, Y. (2012). Genome change in wheat observed through the structure and expression of α/β -gliadin genes. *Funct. Integr. Genomics* 12, 341–355. doi: 10.1007/s10142-012-0269-0
- Kim, C.-Y., Quarsten, H., Bergseng, E., Khosla, C., and Sollid, L. M. (2004). Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. *Proc. Natl. Acad. Sci. U.S.A.* 101, 4175–4179. doi: 10.1073/pnas.0306885101
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. doi: 10.1007/BF01731581
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Li, D., Jin, H., Zhang, K., Wang, Z., Wang, F., Zhao, Y., et al. (2018). Analysis of the Gli-D2 locus identifies a genetic target for simultaneously improving the breadmaking and health-related traits of common wheat. *Plant J.* 95, 414–426. doi: 10.1111/tpj.13956
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, Y., Xin, R., Zhang, D., and Li, S. (2014). Molecular characterization of α -gliadin genes from common wheat cultivar Zhengmai 004 and their role in quality and celiac disease. *Crop. J.* 2, 10–21. doi: 10.1016/j.cj.2013.11.003
- Metakovsky, E., Melnik, V., Rodriguez-Quirjano, M., Upelnik, V., and Carrillo, J. M. (2018). A catalog of gliadin alleles: polymorphism of 20th-century common wheat germplasm. *Crop. J.* 6, 628–641. doi: 10.1016/j.cj.2018.02.003
- Noma, S., Kawaura, K., Hayakawa, K., Abe, C., Tsuge, N., and Ogihara, Y. (2016). Comprehensive molecular characterization of the α/β -gliadin multigene family in hexaploid wheat. *Mol. Genet. Genomics* 291, 65–77. doi: 10.1007/s00438-015-1086-7
- Ozuna, C. V., Iehisa, J. C. M., Giménez, M. J., Alvarez, J. B., Sousa, C., and Barro, F. (2015). Diversification of the celiac disease α -gliadin complex in wheat: a 33-mer peptide with six overlapping epitopes, evolved following polyploidization. *Plant J.* 82, 794–805. doi: 10.1111/tpj.12851
- Paape, T., Briskine, R. V., Halstead-Nussloch, G., Lischer, H. E. L., Shimizu-Inatsugi, R., Hatakeyama, M., et al. (2018). Patterns of polymorphism and selection in the subgenomes of the allopolyploid *Arabidopsis kamchatica*. *Nat. Commun.* 9:3909. doi: 10.1038/s41467-018-06108-1
- Paape, T., Hatakeyama, M., Shimizu-Inatsugi, R., Cereghetti, T., Onda, Y., Kenta, T., et al. (2016). Conserved but attenuated parental gene expression in allopolyploids: constitutive zinc hyperaccumulation in the allotetraploid *arabidopsis kamchatica*. *Mol. Biol. Evol.* 33, 2781–2800. doi: 10.1093/molbev/msw141
- Payne, P. I. (1987). Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. *Ann. Rev. Plant Physiol.* 38, 141–153. doi: 10.1146/annurev.pp.38.060187.001041
- Petersen, J., Kooy-Winkelaar, Y., Loh, K. L., Tran, M., van Bergen, J., Koning, F., et al. (2016). Diverse T Cell receptor gene usage in HLA-DQ8-Associated celiac disease converges into a consensus binding solution. *Structure* 24, 1643–1657. doi: 10.1016/j.str.2016.07.010
- Petersen, J., Montserrat, V., Mujico, J. R., Loh, K. L., Beringer, D. X., van Lummel, M., et al. (2014). T-cell receptor recognition of HLA-DQ2–gliadin complexes associated with celiac disease. *Nat. Struct. Mol. Biol.* 21, 480–488. doi: 10.1038/nsmb.2817
- Pronin, D., Börner, A., and Scherf, K. A. (2021). Old and modern wheat (*Triticum aestivum* L.) cultivars and their potential to elicit celiac disease. *Food Chem.* 339:127952. doi: 10.1016/j.foodchem.2020.127952
- Qi, P.-F., Chen, Q., Ouellet, T., Wang, Z., Le, C.-X., Wei, Y.-M., et al. (2013). The molecular diversity of α -gliadin genes in the tribe Triticeae. *Genetica* 141, 303–310. doi: 10.1007/s10709-013-9729-2
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A., and Mesirov, J. P. (2017). Variant review with the integrative genomics viewer. *Cancer Res.* 77, e31–e34. doi: 10.1158/0008-5472.CAN-17-0337
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Ruiz-Carnicer, A., Comino, I., Segura, V., Ozuna, C. V., Moreno, M. L., López-Casado, M. A., et al. (2019). Celiac immunogenic potential of α -Gliadin epitope variants from triticum and aegilops species. *Nutrients* 11:220. doi: 10.3390/nu11020220
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454
- Salentijn, E. M., Goryunova, S. V., Bas, N., van der Meer, I. M., van den Broeck, H. C., Bastien, T., et al. (2009). Tetraploid and hexaploid wheat varieties reveal large differences in expression of alpha-gliadins from homoeologous Gli-2 loci. *BMC Genomics* 10:48. doi: 10.1186/1471-2164-10-48
- Sánchez-León, S., Gil-Humanes, J., Ozuna, C. V., Giménez, M. J., Sousa, C., Voytas, D. F., et al. (2018). Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnol. J.* 16, 902–910. doi: 10.1111/pbi.12837
- Schaart, J. G., Salentijn, E. M. J., Goryunova, S. V., Chidzanga, C., Esselink, D. G., Gosman, N., et al. (2021). Exploring the alpha-gliadin locus: the 33-mer peptide with six overlapping coeliac disease epitopes in *Triticum aestivum* is derived from a subgroup of *Aegilops tauschii*. *Plant J.* 106, 86–94. doi: 10.1111/tpj.15147
- Scherf, K. A., Koehler, P., and Wieser, H. (2016). Gluten and wheat sensitivities – An overview. *J. Cereal Sci.* 67, 2–11. doi: 10.1016/j.jcs.2015.07.008
- Shan, L., Molberg, O., Parrot, I., Hausch, F., Filiz, F., Gray, G. M., et al. (2002). Structural basis for gluten intolerance in celiac sprue. *Science* 297, 2275–2279. doi: 10.1126/science.1074129
- Shewry, P. R. (2018). Do ancient types of wheat have health benefits compared with modern bread wheat? *J. Cereal Sci.* 79, 469–476. doi: 10.1016/j.jcs.2017.11.010
- Shewry, P. R., and Halford, N. G. (2002). Cereal seed storage proteins: structures, properties and role in grain utilization. *J. Exp. Bot.* 53, 947–958. doi: 10.1093/jxb/53.370.947
- Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M., and Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Sec.* 5, 291–317. doi: 10.1007/s12571-013-0263-y
- Shimizu, K. K., Copetti, D., Okada, M., Wicker, T., Tameshige, T., Hatakeyama, M., et al. (2020). De novo genome assembly of the Japanese wheat cultivar norin 61 highlights functional variation in flowering time and fusarium resistance genes in East Asian Genotypes. *Plant Cell Physiol.* 62, 8–27. doi: 10.1093/pcp/pcaa152
- Singh, P., Arora, A., Strand, T. A., Leffler, D. A., Catassi, C., Green, P. H., et al. (2018). Global prevalence of celiac disease: systematic review and meta-analysis. *Clin. Gastroenterol. Hepatol.* 16, 823.e2–836.e2. doi: 10.1016/j.cgh.2017.06.037
- Sollid, L. M., Qiao, S.-W., Anderson, R. P., Gianfrani, C., and Koning, F. (2012). Nomenclature and listing of celiac disease relevant gluten T-cell epitopes restricted by HLA-DQ molecules. *Immunogenetics* 64, 455–460. doi: 10.1007/s00251-012-0599-z
- Urade, R., Sato, N., and Sugiyama, M. (2018). Gliadins from wheat grain: an overview, from primary structure to nanostructures of aggregates. *Biophys. Rev.* 10, 435–443. doi: 10.1007/s12551-017-0367-2

- van Herpen, T. W., Goryunova, S. V., van der Schoot, J., Mitreva, M., Salentijn, E., Vorst, O., et al. (2006). Alpha-gliadin genes from the A, B, and D genomes of wheat contain different sets of celiac disease epitopes. *BMC Genomics* 7:1. doi: 10.1186/1471-2164-7-1
- Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton, J., et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature* 588, 277–283. doi: 10.1038/s41586-020-2961-x
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A. M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449. doi: 10.1093/genetics/155.1.431
- Yang, Z., Wong, W. S. W., and Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22, 1107–1118. doi: 10.1093/molbev/msi097

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Halstead-Nussloch, Tanaka, Copetti, Paape, Kobayashi, Hatakeyama, Kanamori, Wu, Mascher, Kawaura, Shimizu and Handa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genome-Wide Linkage Mapping for Preharvest Sprouting Resistance in Wheat Using 15K Single-Nucleotide Polymorphism Arrays

OPEN ACCESS

Edited by:

Lee Hickey,
The University of
Queensland, Australia

Reviewed by:

Jirui Wang,
Sichuan Agricultural University, China
Cheng Chang,
Anhui Agricultural University, China

*Correspondence:

Shuanghe Cao
caoshuanghe@caas.cn
Yan Zhang
zhangyan07@caas.cn

†These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 29 July 2021

Accepted: 13 September 2021

Published: 14 October 2021

Citation:

Li L, Zhang Y, Zhang Y, Li M, Xu D,
Tian X, Song J, Luo X, Xie L, Wang D,
He Z, Xia X, Zhang Y and Cao S
(2021) Genome-Wide Linkage
Mapping for Preharvest Sprouting
Resistance in Wheat Using 15K
Single-Nucleotide Polymorphism
Arrays. *Front. Plant Sci.* 12:749206.
doi: 10.3389/fpls.2021.749206

Lingli Li^{1†}, Yingjun Zhang^{2†}, Yong Zhang¹, Ming Li¹, Dengan Xu³, Xiuling Tian¹, Jie Song¹,
Xumei Luo¹, Lina Xie¹, Desen Wang¹, Zhonghu He^{1,4}, Xianchun Xia¹, Yan Zhang^{1*} and
Shuanghe Cao^{1*}

¹ Institute of Crop Sciences, National Wheat Improvement Center, Chinese Academy of Agricultural Sciences, Beijing, China,

² Hebei Laboratory of Crop Genetics and Breeding, Institute of Cereal and Oil Crops, Hebei Academy of Agriculture and
Forestry Sciences, Shijiazhuang, China, ³ Shandong Provincial Key Laboratory of Dryland Farming Technology, College of
Agronomy, Qingdao Agricultural University, Qingdao, China, ⁴ International Maize and Wheat Improvement Center (CIMMYT)
China Office, Beijing, China

Preharvest sprouting (PHS) significantly reduces grain yield and quality. Identification of genetic loci for PHS resistance will facilitate breeding sprouting-resistant wheat cultivars. In this study, we constructed a genetic map comprising 1,702 non-redundant markers in a recombinant inbred line (RIL) population derived from cross Yangxiaomai/Zhongyou9507 using the wheat 15K single-nucleotide polymorphism (SNP) assay. Four quantitative trait loci (QTL) for germination index (GI), a major indicator of PHS, were identified, explaining 4.6–18.5% of the phenotypic variances. Resistance alleles of *Qphs.caas-3AL*, *Qphs.caas-3DL*, and *Qphs.caas-7BL* were from Yangxiaomai, and Zhongyou9507 contributed a resistance allele in *Qphs.caas-4AL*. No epistatic effects were detected among the QTL, and combined resistance alleles significantly increased PHS resistance. Sequencing and linkage mapping showed that *Qphs.caas-3AL* and *Qphs.caas-3DL* corresponded to grain color genes *Tamyb10-A* and *Tamyb10-D*, respectively, whereas *Qphs.caas-4AL* and *Qphs.caas-7BL* were probably new QTL for PHS. We further developed cost-effective, high-throughput kompetitive allele-specific PCR (KASP) markers tightly linked to *Qphs.caas-4AL* and *Qphs.caas-7BL* and validated their association with GI in a test panel of cultivars. The resistance alleles at the *Qphs.caas-4AL* and *Qphs.caas-7BL* loci were present in 72.2 and 16.5% cultivars, respectively, suggesting that the former might be subjected to positive selection in wheat breeding. The findings provide not only genetic resources for PHS resistance but also breeding tools for marker-assisted selection.

Keywords: dormancy, KASP marker, QTL mapping, SNP chip, *Triticum aestivum*

INTRODUCTION

Preharvest sprouting (PHS) refers to the germination of physiologically mature grains in spikes before harvest under rainy weather or humid environment (Groos et al., 2002). PHS is a major problem in cereal production and causes losses in seed vitality, yield, and quality (Xu et al., 2019). Wheat (*Triticum aestivum* L.) is one of the most important staple crops. The average annual loss of wheat caused by PHS exceeds \$1 billion worldwide (Shao et al., 2018). Identification of genetic loci for PHS should be helpful for breeding resistant wheat cultivars.

Preharvest sprouting is a complex trait influenced by genetic and environmental factors (Barrero et al., 2015; Wang et al., 2019). Seed dormancy, an adaptive trait that prevents seeds from germinating, even under favorable conditions, is a major genetic factor for PHS (Née et al., 2017). Germination index (GI) is a common parameter to quantify genetic mechanisms underlying seed dormancy and PHS (Barrero et al., 2015). Some non-dormancy factors, such as spike erectness, spike and awn structure, and openness of florets, are also associated with PHS (Zhu et al., 2019).

Molecular markers have an important role in determining the genetic basis of agronomic traits in wheat (Collard and Mackill, 2008). Markers tightly linked with genes for PHS resistance can be used in marker-assisted selection (MAS). Using diverse mapping populations, many quantitative trait loci (QTL) for PHS resistance or seed dormancy on all 21 wheat chromosomes have been reported (Cao et al., 2016; Lin et al., 2016; Yang et al., 2019; Zhu et al., 2019; Tai et al., 2021). Among them, QTL on group 3 chromosomes and chromosome 4AL have major effects on PHS (Mori et al., 2005; Chen et al., 2008; Ogonnaya et al., 2008; Shao et al., 2018; Vetch et al., 2019). A few genes for PHS in wheat were also isolated by map-based cloning. For example, *TaPHS1*, an *MFT* homolog, is the causal gene in *Qphs.pseru-3AS* (Liu et al., 2013; Jiang et al., 2018; Wang et al., 2020); *Tamyb10* genes at the *R* loci on chromosomes 3A, 3B, and 3D control grain coat color by regulating the accumulation of anthocyanins (Himi and Noda, 2005; Wang et al., 2016; Mares and Himi, 2021); *Mitogen-activated protein kinase kinase 3* (*MKK3*) is the causal gene of *Phs1-4AL* for seed dormancy in wheat (Torada et al., 2016; Martinez et al., 2020); and tandem duplicated plasma membrane protein genes (*PM19*) have been validated as candidates for a major dormancy QTL on chromosome 4AL through transcriptome analysis (Barrero et al., 2015; Shorinola et al., 2016). Homology-based cloning approaches were also used to identify PHS-related genes, such as *TaSdr* (Zhang et al., 2014, 2017), *Vp-1* (McCarty et al., 1991; Yang et al., 2007, 2013; Feng et al., 2017; Zhou et al., 2017), *Qsd1* (Sato et al., 2016; Onishi et al., 2017), and *DOG1* (Ashikawa et al., 2010; Nakabayashi et al., 2012; Rikiishi and Maekawa, 2014; Nishimura et al., 2018).

Yangxiaomai, a red-seeded Chinese landrace, has a high level of PHS resistance, whereas white-seeded Zhongyou9507 with good processing quality is susceptible to PHS. The objectives of this study are to mine QTL for PHS resistance in a recombinant inbred line (RIL) population derived from a Yangxiaomai/Zhongyou9507 cross and to develop breeding-friendly markers for selection of PHS-resistant varieties.

MATERIALS AND METHODS

Plant Materials and Field Trials

The parents Yangxiaomai and Zhongyou9507 and 194 F₆ RILs were planted at Beijing and Shijiazhuang (Hebei Province) in the 2011–2012 cropping season and at Gaoyi (Hebei Province) and Xinxiang (Henan Province) in the 2019–2020 cropping season. Field experiments were arranged in randomized complete blocks with three replications. Each plot was 1 m single row in which 30 seeds were sown. A panel of 101 wheat cultivars (Zhang et al., 2017) was used to determine the genetic effects of the QTL of interest.

Evaluation of PHS Resistance

The GI was used as an indicator of PHS. Five spikes were harvested from each plot at physiological maturity characterized by loss of green color from the spike (Liu et al., 2013). The harvested spikes were air-dried for 2 days at room temperature, hand-threshed to avoid damage to embryos, and then stored in a refrigerator at −20°C to maintain dormancy until phenotyping (Zhang et al., 2017). Seeds were sterilized with 1% (V/V) of NaClO for 20 min, followed by three rinses with sterile water. Notably, 100 healthy seeds of each line were incubated in a 90 mm Petri dish containing a filter paper and 8 ml of distilled water at 20°C for 7 days. Germinated seeds were counted every day and removed. GI was calculated according to the following formula (Walker-Simmons, 1988): $GI = \frac{7 \times n_1 + 6 \times n_2 + 5 \times n_3 + 4 \times n_4 + \dots + 1 \times n_7}{7 \times \text{total grains}} \times 100$, where $n_1, n_2, \dots, n_7$ are the number of seeds germinated on the first, second, and subsequent days until the seventh day.

Statistical Analyses

Phenotypic correlation coefficients among environments, the best linear unbiased prediction (BLUP) values, ANOVA, and *t*-tests were carried out using SAS 9.4 software (SAS Institute Inc., Cary, NC, USA). Broad-sense heritability (H^2) for PHS was calculated using the following formula: $H^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_{ge}^2/e + \sigma_e^2/re)$, where σ_g^2 , σ_{ge}^2 , and σ_e^2 are the variances of genotype, genotype-environment interaction, and residual error, respectively, *r* is the number of replicates, and *e* is the number of environments (Nyquist and Baker, 1991).

Genotyping and Linkage Map Construction

The 194 RILs and parents were genotyped with the wheat 15K single-nucleotide polymorphism (SNP) chips containing 13,947 SNP markers at China Golden Marker (Beijing) Biotech Co., Ltd. (<http://www.cgmb.com.cn/>). To reduce the impact of low-quality SNPs on mapping results, SNP data were processed as follows: (1) Heterozygous loci were treated as missing data, and (2) SNPs with low minor allele frequencies (<0.3) and missing values (>0.2) were excluded using Tassel version 5.0 (Bradbury et al., 2007). Redundant markers were eliminated by the BIN function in QTL IciMapping version 4.2 (Meng et al., 2015). Joinmap version 4.0 was used for linkage map construction (Stam, 1993), and genetic distances between markers were calculated according to the Kosambi mapping function (Kosambi, 1943).

QTL Analysis

Composite interval mapping (CIM) was used to search QTL of phenotypic traits from each environment and BLUP value by Windows QTL Cartographer version 2.5 (Zeng, 1994; Wang et al., 2012). Significant QTLs were identified if the logarithm of odds (LOD) values were more than the threshold of 2.5 (Yan et al., 2006). According to International Wheat Genome Sequencing Consortium (IWGSC) RefSeq 1.0 [International Wheat Genome Sequencing Consortium (IWGSC), 2018] <http://plants.ensembl.org/index.html>, the physical positions of QTL were figured out by the closely linked flanking markers. The genetic maps of QTL were drawn using MapChart version 2.3 (Voorrips, 2002). The analysis of epistatic effects among the QTL was performed using IciMapping version 4.2.

KASP Marker Development and Validation

Kompetitive allele-specific PCR primers (Supplementary Table 5) were designed using PolyMarker (Ramirez-Gonzalez et al., 2015). Primer mixture was prepared with 46 μ l of H₂O, 30 μ l of common primer (100 μ M), and 12 μ l of each tailed primer (100 μ M). PCR was performed in a 384-well plate, and each reaction of $\sim$ 3 μ l comprising 20–30 ng of genomic DNA, 1.5 μ l of 2 $\times$ KASP master mix (V4.0, LGC Genomics, Hoddesdon, UK), 0.0336 μ l of primer mixture, and 1.5 μ l of H₂O. Thermal cycling profile of PCR consisted of hot start at 95°C for 15 min, 10 touchdown cycles (95°C for 20 s and touchdown at 65 and -1° C per cycle for 25 s), and followed by 35 additional cycles (95°C for 20 s and 57°C for 60 s). The 384-well optically clear plates were read on PHERAstarplus SNP (BMG Labtech GmbH, Ortenberg, Germany), and data analysis was carried out using KlusterCaller (LGC, Hoddesdon, UK).

RESULTS

Phenotypic Evaluation

The parents Yangxiaomai and Zhongyou9507 and RILs were evaluated for PHS resistance in four environments. The phenotypes of seed germination in parents Yangxiaomai and Zhongyou9507 were depicted in Supplementary Figure 1. Yangxiaomai had a significantly lower GI (4.3%) than Zhongyou9507 (72.3%) across environments (Supplementary Figure 2). GI for the RIL population showed continuous variation, indicating polygenic inheritance (Supplementary Figure 2). The GI frequencies were skewed toward resistance, suggesting the presence of major genetic loci. GI was significantly correlated among environments with correlation coefficients of 0.53–0.73 (Supplementary Table 1). ANOVA indicated that genotypes and environments, as well as their interactions, had significant effects on GI (Supplementary Table 2). The broad-sense heritability of GI was high (0.88) across environments, denoting that GI variation was mainly determined by genotypes.

Linkage Map Construction and QTL Analysis

The RIL population was genotyped by 15K SNP chips, and 4,515 polymorphic markers were used to construct a

genetic map with 1,702 bin markers, spanning 2,630.9 cM on 21 wheat chromosomes (Supplementary Table 3 and Supplementary Figure 3). The average linkage group was 125.3 cM with an average marker interval of 1.6 cM. Overall, 1,743 (38.6%), 1,750 (38.8%), and 1,022 (22.6%) markers were mapped to the A, B, and D sub-genomes with average marker densities of 1.5, 1.2, and 2.2 cM, respectively (Supplementary Table 3 and Supplementary Figure 3).

Four QTLs for PHS were detected by CIM on the linkage groups 3AL (*Qphs.caas-3AL*), 3DL (*Qphs.caas-3DL*), 4AL (*Qphs.caas-4AL*), and 7BL (*Qphs.caas-7BL*) (Table 1 and Figure 1). Alleles for resistance to PHS on chromosome arms 3AL, 3DL, and 7BL loci were from Yangxiaomai, whereas the resistance allele on 4AL was contributed by Zhongyou9507. *Qphs.caas-3DL* was identified across all four environments and explained 8.9–18.5% of the phenotypic variances; *Qphs.caas-3AL* and *Qphs.caas-4AL* were detected in three of four environments, explaining 10.5–13.5 and 4.6–10.6% of the phenotypic variances, respectively; and *Qphs.caas-7BL* accounting for 5.0–6.7% of the phenotypic variances was detected in two environments.

Combinational Effects of the Stable QTL for PHS Resistance

Quantitative trait loci for a given trait detected in more than one-half of tested environments can be considered stable genetic loci (Cao et al., 2020). The QTL *Qphs.caas-3AL*, *Qphs.caas-3DL*, and *Qphs.caas-4AL* fulfilled that criterion. To confirm their genetic effects on PHS, the population was classified into eight groups based on the closest flanking SNPs for each QTL (Supplementary Table 4). *Qphs.caas-3AL*, *Qphs.caas-3DL*, and *Qphs.caas-4AL* were temporarily designated as the loci 1, 2, and 3, respectively, and R and S represented resistance and susceptible alleles, respectively. The GI values of eight groups (i.e., $I^R2^R3^R$, $I^R2^R3^S$, $I^R2^S3^R$, $I^R2^S3^S$, $I^S2^R3^R$, $I^S2^R3^S$, $I^S2^S3^R$, and $I^S2^S3^S$) were compared across four environments (Figure 2 and Supplementary Table 4).

The eight groups were ranked according to the GI in four environments and BLUP values. In general, more resistance alleles conferred lower GI, demonstrating cumulative effects of resistance alleles at the three loci (Figure 2). RILs with genotype $I^R2^R3^R$ had the lowest GI, and those with $I^S2^S3^S$ exhibited the highest GI across all environments. However, the GI of RILs with $I^R2^R3^S$ were higher than those with $I^S2^R3^S$ in Beijing 2012, suggesting that the genetic effect of locus 1, *Qphs.caas-3AL*, was significantly affected by the environment in some cases. No epistatic effects were detected among the QTL. Regression analysis also showed that the lines carrying more resistance alleles had higher PHS resistance in individual environments and the BLUP data (Supplementary Figure 4). Thus, the pyramiding of resistance alleles was effective in improving PHS resistance.

Relationship of *Qphs.caas-3AL* and *Qphs.caas-3DL* With Tamyb10

According to IWGSC RefSeq 1.0, *Qphs.caas-3AL* and *Qphs.caas-3DL* were delimited in the intervals of 700.4–709.2 Mb and

TABLE 1 | QTL for GI detected in the Yangxiaomai/Zhongyou9507 RIL population.

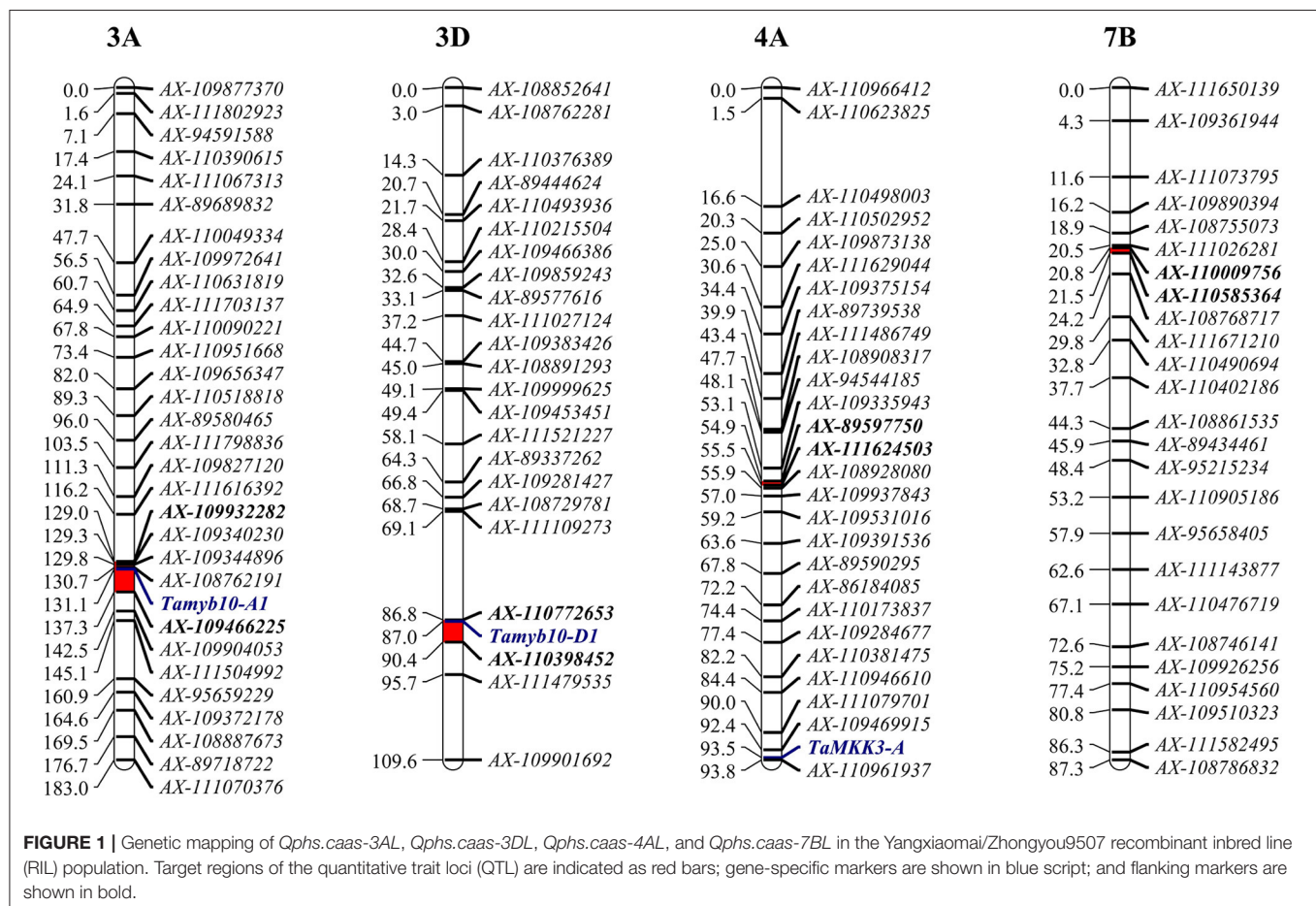
QTL	Environment	Flanking marker	Physical interval (Mb)	LOD ^a	PVE (%) ^b	Add ^c
<i>Qphs.caas-3AL</i>	2020 XX	AX-109932282-AX-109340230	700.4–702.5	6.3	10.5	4.7
	2020 GY	AX-108762191-AX-109466225	703.9–709.2	7.5	13.5	6.7
	2012 SJZ	AX-109932282-AX-109340230	700.4–702.5	6.5	11.2	5.4
	BLUP	AX-109932282-AX-109340230	700.4–702.5	6.8	10.6	3.6
<i>Qphs.caas-3DL</i>	2020 XX	AX-110772653-AX-110398452	570.2–575.1	7.1	12.2	5.1
	2020 GY	AX-110772653-AX-110398452	570.2–575.1	10.8	18.5	7.9
	2012 SJZ	AX-110772653-AX-110398452	570.2–575.1	7.3	13.4	6.1
	2012 BJ	AX-110772653-AX-110398452	570.2–575.1	4.9	8.9	6.6
<i>Qphs.caas-4AL</i>	BLUP	AX-110772653-AX-110398452	570.2–575.1	8.7	14.4	4.2
	2020 XX	AX-89597750-AX-111624503	489.1–532.2	3.3	5.2	–3.3
	2020 GY	AX-89597750-AX-111624503	489.1–532.2	3.0	4.6	–3.8
	2012 BJ	AX-89597750-AX-111624503	489.1–532.2	6.1	10.6	–7.1
<i>Qphs.caas-7BL</i>	BLUP	AX-89597750-AX-111624503	489.1–532.2	5.4	8.2	–3.2
	2020 GY	AX-110009756-AX-110585364	522.6–529.7	3.7	5.7	4.3
	2012 BJ	AX-110009756-AX-110585364	522.6–529.7	3.5	6.1	5.4
	BLUP	AX-110009756-AX-110585364	522.6–529.7	3.2	5.0	2.4

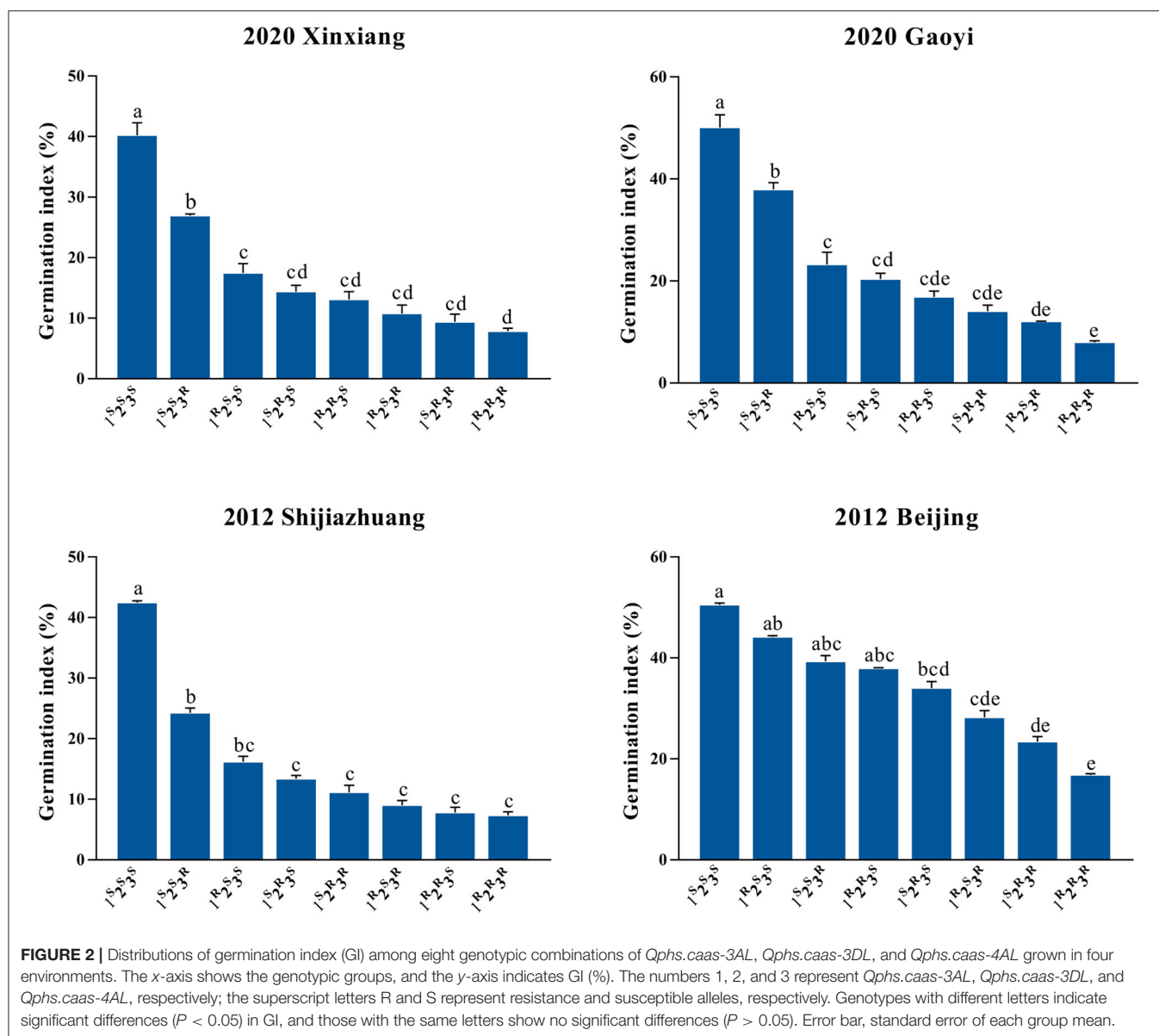
^aA LOD threshold of 2.5 was used for the declaration of QTL.

^bPercentage of phenotypic variances explained by QTL.

^cPositive and negative additive effects indicated increasing effects from Yangxiaomai and Zhongyou9507, respectively.

QTL, quantitative trait loci; GI, germination index; RIL, recombinant inbred line; XX, Xinxiang; GY, Gaoyi; SJZ, Shijiazhuang; BJ, Beijing; LOD, logarithm of odds.





570.2–575.1 Mb, respectively, based on their flanking markers. This placed PHS-related genes *Tamyb10-A1* (~703.9 Mb) and *Tamyb10-D1* (~570.8 Mb), in the regions of *Qphs.caas-3AL* and *Qphs.caas-3DL*, respectively (Himi et al., 2011). To confirm the genetic relationship of the two genes with *Qphs.caas-3AL* and *Qphs.caas-3DL*, we re-genotyped the Yangxiaomai/Zhongyou9507 RIL population using a KASP marker for *Tamyb10-A1* and an STS marker for *Tamyb10-D1* (Supplementary Table 5) (Himi et al., 2011; Wang et al., 2016). In these analyses, *Tamyb10-A1* and *Tamyb10-D1* were mapped to the genetic regions of *Qphs.caas-3AL* and *Qphs.caas-3DL*, respectively (Figure 1). Thus, *Tamyb10-A1* and *Tamyb10-D1* were likely causal genes in *Qphs.caas-3AL* and *Qphs.caas-3DL*, respectively.

Relationship Between *Qphs.caas-4AL* and Reported PHS Resistance Genes on Chromosome 4AL

TaMKK3-A was reported as a major gene controlling seed dormancy on chromosome 4AL (Torada et al., 2016). Based on the IWGSC RefSeq 1.0, *TaMKK3-A* is located at the site of ~609.1 Mb (GenBank accession number: LC091368.1) (Liton et al., 2021). *Qphs.caas-4AL* spans the interval of 489.1–532.2 Mb according to flanking markers AX-89597750 and AX-111624503, suggesting that *TaMKK3-A* is not in the target region of *Qphs.caas-4AL*. We sequenced the *TaMKK3-A* gene in Yangxiaomai and Zhongyou9507 to confirm the relationship between *TaMKK3-A* and *Qphs.caas-4AL*. Sequence analysis showed that *TaMKK3-A* had no polymorphic sites between

TABLE 2 | The effects of *Qphs.caas-4AL* and *Qphs.caas-7BL* on GI in a natural population.

QTL	Genotype	Number	GI (%)	P-value
<i>Qphs.caas-4AL</i>	AA	70	31.82	0.03*
	BB	27	40.95	
<i>Qphs.caas-7BL</i>	AA	81	36.73	0.008**
	BB	16	23.47	

* $P < 0.05$; ** $P < 0.01$.

The wheat lines in the natural population are listed in **Supplementary Table 8**.

GI, germination index; QTL, quantitative trait loci; AA, the allele from Zhongyou9507; BB, the allele from Yangxiaomai.

two parents in all exons, but three SNPs were detected in the introns (**Supplementary Figure 5**). A KASP marker KASP-6464 was developed for *TaMKK3-A*. Linkage mapping showed that KASP-6464 was out of the target region of *Qphs.caas-4AL* (**Figure 1**). Association analysis also indicated that *TaMKK3-A* had no significant effect on GI in three environments (**Supplementary Table 6**). Therefore, *TaMKK3-A* was not a candidate gene in *Qphs.caas-4AL*.

PM19-A1 is a second PHS-related gene in chromosome 4AL. However, it is positioned at ~604.1 Mb, which is far from *Qphs.caas-4AL* (489.1–532.2 Mb) according to the IWGSC RefSeq 1.0. We also compared the sequences of *PM19-A1* between Yangxiaomai and Zhongyou9507 and found no polymorphic sites in the open reading frame and the previously reported 18 bp indel in the promoter (Barrero et al., 2015; Shorinola et al., 2016, 2017). Therefore, *PM19-A1* was not the causal gene in *Qphs.caas-4AL*.

No genes related to PHS were isolated on 7B so far, so no candidate genes could be used to perform sequencing and genetic mapping analyses for *Qphs.caas-7BL*. *Qphs.caas-7BL* is defined in the interval of 522.6–529.7 Mb based on its flanking markers according to IWGSC RefSeq 1.0. QTL related to PHS, which have been reported on 7B, were summarized in **Supplementary Table 7**. The location of *Qphs.caas-7BL* is different from those of the previously reported QTL based on their flanking markers.

Genetic Effects of *Qphs.caas-4AL* and *Qphs.caas-7BL* on GI in a Panel of Cultivars

The causal genes of *Qphs.caas-4AL* and *Qphs.caas-7BL* remained unknown although they could explain 4.6–10.6% of the phenotypic variances. To further decipher the importance of *Qphs.caas-4AL* and *Qphs.caas-7BL*, we attempted to identify their genetic effects in a panel of cultivars. An SNP AX-89597750 closely linked with *Qphs.caas-4AL* was converted to a KASP marker. The KASP marker was mapped at the position of AX-89597750, indicating that it could act as a closely linked marker of *Qphs.caas-4AL*. We detected the genetic effect of *Qphs.caas-4AL* on GI using the KASP marker in the cultivar panel (**Supplementary Table 8**). Genotypes with the resistance allele had significantly lower GI than those with the susceptible allele (**Table 2**). Moreover, a majority of genotypes (72.2%) possessed the resistance allele of *Qphs.caas-4AL*, suggesting that it had

been subjected to selection in wheat breeding (**Table 2**). Another KASP marker was developed based on SNP AX-9496498 closely linked to *Qphs.caas-7BL*. The KASP marker was mapped to the target region of *Qphs.caas-7BL* in the mapping population. Like *Qphs.caas-4AL*, *Qphs.caas-7BL* was also significantly associated with PHS resistance (**Table 2**), but only 16.5% of cultivars carried the resistance allele.

DISCUSSION

Grain Color Is a Major Factor Modulating PHS

The red-grain Yangxiaomai and white-grain Zhongyou9507 have a relatively low and higher GI, respectively. In this study, we confirmed that *Qphs.caas-3AL* and *Qphs.caas-3DL* co-localized with *Tamyb10-A1* and *Tamyb10-D1*, respectively, at the *R* loci for grain color (Himi et al., 2011; Lang et al., 2021; Mares and Himi, 2021). Wang et al. (2016) also observed that *Tamyb10-D1* was significantly ($P < 0.001$) associated with GI in a natural population. Thus, grain color is probably regulated by *Tamyb10* alleles in *Qphs.caas-3AL* and *Qphs.caas-3DL*, which also cause significant differences in GI between Yangxiaomai and Zhongyou9507. Pleiotropic QTL for grain color and PHS resistance on chromosomes 3AL and 3DL were identified in a genome-wide association study conducted by Lin et al. (2016). These findings also confirm that grain color has a great effect on PHS resistance in wheat breeding.

Qphs.caas-4AL Has Potential Value for Wheat Breeding

Qphs.caas-4AL, as a stable QTL, accounted for 4.6–10.6% of the phenotypic variances. Although quite a few QTL for wheat PHS resistance on chromosome 4A have also been reported (Kato et al., 2001; Mares et al., 2005; Torada et al., 2005; Imtiaz et al., 2008; Rasul et al., 2009; Singh et al., 2010; Kulwal et al., 2012; Cabral et al., 2014; Jiang et al., 2015; Cao et al., 2016; Zhou et al., 2017; Martinez et al., 2018; Zuo et al., 2019; Liton et al., 2021), *Qphs.caas-4AL* appears to be unique based on genetic mapping and physical locations of the flanking SNPs according to IWGSC RefSeq 1.0 (**Supplementary Table 9**). QTL pyramiding plays an important role in breeding, and resistance allele combinations of QTL for PHS have been reported previously (Imtiaz et al., 2008; Shao et al., 2018; Liton et al., 2021). In

this study, we identified that the RILs combining resistance alleles in *Qphs.caas-3AL*, *Qphs.caas-3DL*, and *Qphs.caas-4AL* had the lowest GI (Figure 2). *Qphs.caas-4AL* also improved resistance to PHS in the absence of the alleles for red-grain color (Figure 2 and Supplementary Table 4). This indicated that *Qphs.caas-4AL* could function independently of grain color. We converted an SNP tightly linked to *Qphs.caas-4AL* into a KASP marker. Genotyping by the KASP marker showed that most of the cultivars (72.2%) carried the resistance allele in *Qphs.caas-4AL* in the test panel (Table 2), indicating that the resistance allele of *Qphs.caas-4AL* might undergo positive selection in breeding programs. *Qphs.caas-4AL* is also significantly associated with GI (Table 2). Thus, the KASP marker will be useful for MAS to improve PHS tolerance in wheat. Grain color is an important parameter for wheat appearance quality. Red-grain wheat usually has high resistance to PHS but is adverse to make Chinese traditional food, such as steamed bread and noodles (Fakthongphan et al., 2016; Shao et al., 2018). Thus, *Qphs.caas-4AL* is a better choice for improvement of PHS than *Qphs.caas-3AL* and *Qphs.caas-3DL* at least in Chinese wheat breeding. Overall, these findings indicate that *Qphs.caas-4AL* is a valuable genetic locus for PHS in wheat breeding.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

LL and SC wrote the manuscript. LL, YiZ, ML, DX, XT, JS, and XL performed the experiments. SC and LL analyzed the data. YoZ, LX, and DW participated in the field trials. SC and YaZ designed the experiments. XX and ZH assisted in writing the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was funded by the National Natural Science Foundation of China (Grant Nos. 91935304 and 31571663), the National Key Research and Development Programs of China (Grant No. 2016YFD0100502), and the CAAS Science and Technology Innovation Program.

REFERENCES

- Ashikawa, I., Abe, F., and Nakamura, S. (2010). Ectopic expression of wheat and barley *DOG1*-like genes promotes seed dormancy in *Arabidopsis*. *Plant Sci.* 179, 536–542. doi: 10.1016/j.plantsci.2010.08.002
- Barrero, J. M., Cavanagh, C., Verbyla, K. L., Tibbits, J. F., Verbyla, A. P., Huang, B. E., et al. (2015). Transcriptomic analysis of wheat near-isogenic lines identifies *PM19-A1* and *A2* as candidates for a major dormancy QTL. *Genome Biol.* 16:93. doi: 10.1186/s13059-015-0665-6
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping

ACKNOWLEDGMENTS

The authors thank Prof. R. A. McIntosh, Plant Breeding Institute, University of Sydney, for critical review of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.749206/full#supplementary-material>

Supplementary Figure 1 | Seed germination of Yangxiaomai and Zhongyou9507. YXM, Yangxiaomai; ZY9507, Zhongyou9507.

Supplementary Figure 2 | Distributions of GI in the Yangxiaomai/Zhongyou9507 RIL population in four environments. Black arrows point to the GI of the parents. GI, germination index; RIL, recombinant inbred line; YXM, Yangxiaomai; ZY9507, Zhongyou9507.

Supplementary Figure 3 | Genetic map of the Yangxiaomai/Zhongyou9507 RIL population determined from 15K SNP arrays. RIL, recombinant inbred line; SNP, single-nucleotide polymorphism.

Supplementary Figure 4 | Linear regression between the resistance allele number of the QTL and GI in four environments and BLUP values in the Yangxiaomai/Zhongyou9507 population. The QTL include *Qphs.caas-3AL*, *Qphs.caas-3DL*, *Qphs.caas-4AL*, and *Qphs.caas-7BL*. QTL, quantitative trait loci; GI, germination index; BLUP, best linear unbiased prediction.

Supplementary Figure 5 | Schematic of the open reading frame of *TaMKK3-A* and its polymorphic sites between Yangxiaomai (YXM) and Zhongyou9507 (ZY9507). Bold bars and thin lines indicate exons and introns, respectively.

Supplementary Table 1 | Correlation coefficients of GI from the Yangxiaomai/Zhongyou9507 RIL population among four environments.

Supplementary Table 2 | Variance analysis for GI of the Yangxiaomai/Zhongyou9507 RIL population.

Supplementary Table 3 | Information of genetic map construction for Yangxiaomai/Zhongyou9507 RIL population using 15K SNP chips.

Supplementary Table 4 | GI of eight genotypic groups defined by three stable QTL in the Yangxiaomai/Zhongyou9507 RIL population.

Supplementary Table 5 | Primers used in this study.

Supplementary Table 6 | Association analysis of *TaMKK3-A* and GI in the Yangxiaomai/Zhongyou9507 RIL population.

Supplementary Table 7 | Information of reported QTL for PHS on the chromosome 7B.

Supplementary Table 8 | Genotypes of *Qphs.caas-4AL* and *Qphs.caas-7BL* and GI of 101 wheat cultivars.

Supplementary Table 9 | Information of reported QTL for PHS on the chromosome 4A.

- of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Cabral, A. L., Jordan, M. C., McCartney, C. A., You, F. M., Humphreys, D. G., MacLachlan, R., et al. (2014). Identification of candidate genes, regions and markers for pre-harvest sprouting resistance in wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 14:340. doi: 10.1186/s12870-014-0340-1
- Cao, L., Hayashi, K., Tokui, M., Mori, M., Miura, H., and Onishi, K. (2016). Detection of QTLs for traits associated with pre-harvest sprouting resistance in bread wheat (*Triticum aestivum* L.). *Breed. Sci.* 66, 260–270. doi: 10.1270/jsbbs.66.260

- Cao, S., Xu, D., Hanif, M., Xia, X., and He, Z. (2020). Genetic architecture underpinning yield component traits in wheat. *Theor. Appl. Genet.* 133, 1811–1823. doi: 10.1007/s00122-020-03562-8
- Chen, C., Cai, S., and Bai, G. (2008). A major QTL controlling seed dormancy and pre-harvest sprouting resistance on chromosome 4A in a Chinese wheat landrace. *Mol. Breed.* 21, 351–358. doi: 10.1007/s11032-007-9135-5
- Collard, B. C., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Phil. Trans. R. Soc. B.* 363, 557–572. doi: 10.1098/rstb.2007.2170
- Fakhthongphan, J., Bai, G., St Amand, P., Graybosch, R. A., and Baenziger, P. S. (2016). Identification of markers linked to genes for sprouting tolerance (independent of grain color) in hard white winter wheat (HWWW). *Theor. Appl. Genet.* 129, 419–430. doi: 10.1007/s00122-015-2636-4
- Feng, Y., Qu, R., Liu, S., and Yang, Y. (2017). Rich haplotypes of *Viviparous-1* in *Triticum aestivum* subsp. *spelta* with different abscisic acid sensitivities. *J. Sci. Food Agric.* 97, 497–504. doi: 10.1002/jsfa.7751
- Groos, C., Gay, G., Perretant, M. R., Gervais, L., Bernard, M., Dedryver, F., et al. (2002). Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white×red grain bread-wheat cross. *Theor. Appl. Genet.* 104, 39–47. doi: 10.1007/s001220200004
- Himi, E., Maekawa, M., Miura, H., and Noda, K. (2011). Development of PCR markers for *Tamyb10* related to R-1, red grain color gene in wheat. *Theor. Appl. Genet.* 122, 1561–1576. doi: 10.1007/s00122-011-1555-2
- Himi, E., and Noda, K. (2005). Red grain colour gene (*R*) of wheat is a Myb-type transcription factor. *Euphytica* 143, 239–242. doi: 10.1007/s10681-005-7854-4
- Imtiaz, M., Ogbonnaya, F. C., Oman, J., and van Ginkel, M. (2008). Characterization of quantitative trait loci controlling genetic variation for preharvest sprouting in synthetic backcross-derived wheat lines. *Genetics* 178, 1725–1736. doi: 10.1534/genetics.107.084939
- International Wheat Genome Sequencing Consortium (IWGSC) (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191
- Jiang, H., Zhao, L., Chen, X., Cao, J., Wu, Z., Liu, K., et al. (2018). A novel 33-bp insertion in the promoter of *TaMFT-3A* is associated with pre-harvest sprouting resistance in common wheat. *Mol. Breed.* 38:69. doi: 10.1007/s11032-018-0830-1
- Jiang, Y., Wang, J., Luo, W., Wei, Y., Qi, P., Liu, Y., et al. (2015). Quantitative trait locus mapping for seed dormancy in different post-ripening stages in a Tibetan semi-wild wheat (*Triticum aestivum* ssp. *tibetanum* Shao). *Euphytica* 203, 557–567. doi: 10.1007/s10681-014-1266-2
- Kato, K., Nakamura, W., Tabiki, T., Miura, H., and Sawada, S. (2001). Detection of loci controlling seed dormancy on group 4 chromosomes of wheat and comparative mapping with rice and barley genomes. *Theor. Appl. Genet.* 102, 980–985. doi: 10.1007/s001220000494
- Kosambi, D. D. (1943). The estimation of map distances from recombination values. *Ann. Eugen.* 12, 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x
- Kulwal, P., Ishikawa, G., Benscher, D., Feng, Z., Yu, L. X., Jadhav, A., et al. (2012). Association mapping for pre-harvest sprouting resistance in white winter wheat. *Theor. Appl. Genet.* 125, 793–805. doi: 10.1007/s00122-012-1872-0
- Lang, J., Fu, Y., Zhou, Y., Cheng, M., Deng, M., Li, M., et al. (2021). *Myb10-D* confers *PHS-3D* resistance to pre-harvest sprouting by regulating NCED in ABA biosynthesis pathway of wheat. *New Phytol.* 230, 1940–1952. doi: 10.1111/nph.17312
- Lin, M., Zhang, D., Liu, S., Zhang, G., Yu, J., Fritz, A. K., et al. (2016). Genome-wide association analysis on pre-harvest sprouting resistance and grain color in U.S. winter wheat. *BMC Genom.* 17:794. doi: 10.1186/s12864-016-3148-6
- Liton, M., McCartney, C. A., Hiebert, C. W., Kumar, S., Jordan, M. C., and Ayele, B. T. (2021). Identification of loci for pre-harvest sprouting resistance in the highly dormant spring wheat RL4137. *Theor. Appl. Genet.* 134, 113–124. doi: 10.1007/s00122-020-03685-y
- Liu, S., Sehgal, S. K., Li, J., Lin, M., Trick, H. N., Yu, J., et al. (2013). Cloning and characterization of a critical regulator for preharvest sprouting in wheat. *Genetics* 195, 263–273. doi: 10.1534/genetics.113.152330
- Mares, D., and Himi, E. (2021). The role of *TaMYB10-A1* of wheat (*Triticum aestivum* L.) in determining grain coat colour and dormancy phenotype. *Euphytica* 217:89. doi: 10.1007/s10681-021-02826-8
- Mares, D., Mrva, K., Cheong, J., Williams, K., Watson, B., Storlie, E., et al. (2005). A QTL located on chromosome 4A associated with dormancy in white-and red-grained wheats of diverse origin. *Theor. Appl. Genet.* 111, 1357–1364. doi: 10.1007/s00122-005-0065-5
- Martinez, S. A., Godoy, J., Huang, M., Zhang, Z., Carter, A. H., Campbell, K. A., et al. (2018). Genome-wide association mapping for tolerance to preharvest sprouting and low falling numbers in wheat. *Front. Plant Sci.* 9:141. doi: 10.3389/fpls.2018.00141
- Martinez, S. A., Shorinola, O., Conselman, S., See, D., Skinner, D. Z., Uauy, C., et al. (2020). Exome sequencing of bulked segregants identified a novel *TaMKK3-A* allele linked to the wheat *ERA8* ABA-hypersensitive germination phenotype. *Theor. Appl. Genet.* 133, 719–736. doi: 10.1007/s00122-019-03503-0
- McCarty, D. R., Hattori, T., Carson, C. B., Vasil, V., Lazar, M., and Vasil, I. K. (1991). The *Viviparous-1* developmental gene of maize encodes a novel transcriptional activator. *Cell* 66, 895–905. doi: 10.1016/0092-8674(91)90436-3
- Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001
- Mori, M., Uchino, N., Chono, M., Kato, K., and Miura, H. (2005). Mapping QTLs for grain dormancy on wheat chromosome 3A and the group 4 chromosomes, and their combined effect. *Theor. Appl. Genet.* 110, 1315–1323. doi: 10.1007/s00122-005-1972-1
- Nakabayashi, K., Bartsch, M., Xiang, Y., Miatton, E., Pellengahr, S., and Yano, R., et al. (2012). The time required for dormancy release in *Arabidopsis* is determined by DELAY OF GERMINATION1 protein levels in freshly harvested seeds. *Plant Cell* 24, 2826–2838. doi: 10.1105/tpc.112.100214
- Née, G., Xiang, Y., and Soppe, W. J. (2017). The release of dormancy, a wake-up call for seeds to germinate. *Curr. Opin. Plant Biol.* 35, 8–14. doi: 10.1016/j.pbi.2016.09.002
- Nishimura, N., Tsuchiya, W., Moresco, J. J., Hayashi, Y., Satoh, K., Kaiwa, N., et al. (2018). Control of seed dormancy and germination by DOG1-AHG1 PP2C phosphatase complex via binding to heme. *Nat. Commun.* 9:2132. doi: 10.1038/s41467-018-04437-9
- Nyquist, W. E., and Baker, R. J. (1991). Estimation of heritability and prediction of selection response in plant populations. *Crit. Rev. Plant Sci.* 10, 235–322. doi: 10.1080/07352689109382313
- Ogbonnaya, F. C., Imtiaz, M., Ye, G., Hearnden, P. R., Hernandez, E., Eastwood, R. F., et al. (2008). Genetic and QTL analyses of seed dormancy and preharvest sprouting resistance in the wheat germplasm CN10955. *Theor. Appl. Genet.* 116, 891–902. doi: 10.1007/s00122-008-0712-8
- Onishi, K., Yamane, M., Yamaji, N., Tokui, M., Kanamori, H., Wu, J., et al. (2017). Sequence differences in the seed dormancy gene *Qsd1* among various wheat genomes. *BMC Genom.* 18:497. doi: 10.1186/s12864-017-4046-2
- Ramirez-Gonzalez, R. H., Uauy, C., and Caccamo, M. (2015). Poly-Marker: a fast polyploid primer design pipeline. *Bioinformatics* 31:2038–2039. doi: 10.1093/bioinformatics/btv069
- Rasul, G., Humphreys, D. G., Brûlé-Babel, A., McCartney, C. A., and Knox, R. M., Somers, D. J. (2009). Mapping QTLs for pre-harvest sprouting traits in the spring wheat cross 'RL4452/AC domain'. *Euphytica* 168, 363–378. doi: 10.1007/s10681-009-9934-3
- Rikiishi, K., and Maekawa, M. (2014). Seed maturation regulators are related to the control of seed dormancy in wheat (*Triticum aestivum* L.). *PLoS ONE* 9:e107618. doi: 10.1371/journal.pone.0107618
- Sato, K., Yamane, M., Yamaji, N., Kanamori, H., Tagiri, A., Schwerdt, J. G., et al. (2016). Alanine aminotransferase controls seed dormancy in barley. *Nat. Commun.* 7:11625. doi: 10.1038/ncomms11625
- Shao, M., Bai, G., Rife, T. W., Poland, J., Lin, M., Liu, S., et al. (2018). QTL mapping of pre-harvest sprouting resistance in a white wheat cultivar Danby. *Theor. Appl. Genet.* 131, 1683–1697. doi: 10.1007/s00122-018-3107-5
- Shorinola, O., Balcáková, B., Hyles, J., Tibbits, J., Hayden, M. J., Holuřová, K., et al. (2017). Haplotype analysis of the pre-harvest sprouting resistance locus *Phs-A1* reveals a causal role of *TaMKK3-A* in global germplasm. *Front. Plant Sci.* 8:1555. doi: 10.3389/fpls.2017.01555
- Shorinola, O., Bird, N., Simmonds, J., Berry, S., Henriksson, T., Jack, P., et al. (2016). The wheat *Phs-A1* pre-harvest sprouting resistance locus delays the rate of seed dormancy loss and maps 0.3 cM distal to the *PM19* genes in UK germplasm. *J. Exp. Bot.* 67, 4169–4178. doi: 10.1093/jxb/erw194

- Singh, R., Matus-Cádiz, M., Båga, M., Hucl, P., and Chibbar, R. N. (2010). Identification of genomic regions associated with seed dormancy in white-grained wheat. *Euphytica* 174, 391–408. doi: 10.1007/s10681-010-0137-8
- Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. *Plant J.* 3, 739–744. doi: 10.1111/j.1365-313X.1993.00739.x
- Tai, L., Wang, H., Xu, X., Sun, W., Chen, K., Ju, L., et al. (2021). Cereal pre-harvest sprouting: a global agricultural disaster regulated by complex genetic and biochemical mechanisms. *J. Exp. Bot.* 72, 2857–2876. doi: 10.1093/jxb/erab024
- Torada, A., Ikeguchi, S., and Koike, M. (2005). Mapping and validation of PCR-based markers associated with a major QTL for seed dormancy in wheat. *Euphytica* 143, 251–255. doi: 10.1007/s10681-005-7872-2
- Torada, A., Koike, M., Ogawa, T., Takenouchi, Y., Tadamura, K., Wu, J., et al. (2016). A causal gene for seed dormancy on wheat chromosome 4A encodes a MAP kinase kinase. *Curr. Biol.* 26, 782–787. doi: 10.1016/j.cub.2016.01.063
- Vetch, J. M., Stougaard, R. N., Martin, J. M., and Giroux, M. J. (2019). Revealing the genetic mechanisms of pre-harvest sprouting in hexaploid wheat (*Triticum aestivum* L.). *Plant Sci.* 281, 180–185. doi: 10.1016/j.plantsci.2019.01.004
- Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77
- Walker-Simmons, M. (1988). Enhancement of ABA responsiveness in wheat embryos by high temperature. *Plant Cell Environ.* 11, 769–775. doi: 10.1111/j.1365-3040.1988.tb01161.x
- Wang, D., Pang, Y., Dong, L., Li, A., Kong, L., and Liu, S. (2020). Allelic impacts on pre-harvest sprouting resistance and favorable haplotypes in *TaPHS1* of Chinese wheat accessions. *Crop J.* 8, 515–521. doi: 10.1016/j.cj.2019.12.003
- Wang, S., Basten, C., and Zeng, Z. (2012). *Windows QTL Cartographer 2.5*. Department of Statistics, North Carolina State University, Raleigh, NC. Available online at: <http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>
- Wang, X., Liu, H., Liu, G., Mia, M. S., Siddique, K., and Yan, G. (2019). Phenotypic and genotypic characterization of near-isogenic lines targeting a major 4BL QTL responsible for pre-harvest sprouting in wheat. *BMC Plant Biol.* 19:348. doi: 10.1186/s12870-019-1961-1
- Wang, Y., Wang, X., Meng, J., Zhang, Y., He, Z., and Yang, Y. (2016). Characterization of *Tamyb10* allelic variants and development of STS marker for pre-harvest sprouting resistance in Chinese bread wheat. *Mol. Breed.* 36:148. doi: 10.1007/s11032-016-0573-9
- Xu, F., Tang, J., Gao, S., Cheng, X., Du, L., and Chu, C. (2019). Control of rice pre-harvest sprouting by glutaredoxin-mediated abscisic acid signaling. *Plant J.* 100, 1036–1051. doi: 10.1111/tpj.14501
- Yan, J., Tang, H., Huang, Y., and Zheng, Y., Li, J. (2006). Quantitative trait loci mapping and epistatic analysis for grain yield and yield components using molecular markers with an elite maize hybrid. *Euphytica* 149, 121–131. doi: 10.1007/s10681-005-9060-9
- Yang, J., Tan, C., Lang, J., Tang, H., Hao, M., and Tan, Z., et al. (2019). Identification of *qPHS.sicau-1B* and *qPHS.sicau-3D* from synthetic wheat for pre-harvest sprouting resistance wheat improvement. *Mol. Breed.* 39:132. doi: 10.1007/s11032-019-1029-9
- Yang, Y., Ma, Y., Xu, Z., Chen, X., He, Z., Yu, Z., et al. (2007). Isolation and characterization of *Viviparous-1* genes in wheat cultivars with distinct ABA sensitivity and pre-harvest sprouting tolerance. *J. Exp. Bot.* 58, 2863–2871. doi: 10.1093/jxb/erm073
- Yang, Y., Zhang, C., Liu, S., Sun, Y., Meng, J., and Xia, L. (2013). Characterization of the rich haplotypes of *Viviparous-1A* in Chinese wheats and development of a novel sequence-tagged site marker for pre-harvest sprouting resistance. *Mol. Breed.* 33, 75–88. doi: 10.1007/s11032-013-9935-8
- Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136, 1457–1468. doi: 10.1093/genetics/136.4.1457
- Zhang, Y., Miao, X., Xia, X., and He, Z. (2014). Cloning of seed dormancy genes (*TaSdr*) associated with tolerance to pre-harvest sprouting in common wheat and development of a functional marker. *Theor. Appl. Genet.* 127, 855–866. doi: 10.1007/s00122-014-2262-6
- Zhang, Y., Xia, X., and He, Z. (2017). The seed dormancy allele *TaSdr-A1a* associated with pre-harvest sprouting tolerance is mainly present in Chinese wheat landraces. *Theor. Appl. Genet.* 130, 81–89. doi: 10.1007/s00122-016-2793-0
- Zhou, Y., Tang, H., Cheng, M. P., Dankwa, K. O., Chen, Z. X., Li, Z. Y., et al. (2017). Genome-wide association study for pre-harvest sprouting resistance in a large germplasm collection of Chinese wheat landraces. *Front. Plant Sci.* 8:401. doi: 10.3389/fpls.2017.00401
- Zhu, Y., Wang, S., Wei, W., Xie, H., Liu, K., Zhang, C., et al. (2019). Genome-wide association study of pre-harvest sprouting tolerance using a 90K SNP array in common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 132, 2947–2963. doi: 10.1007/s00122-019-03398-x
- Zuo, J., Lin, C. T., Cao, H., Chen, F., Liu, Y., and Liu, J. (2019). Genome-wide association study and quantitative trait loci mapping of seed dormancy in common wheat (*Triticum aestivum* L.). *Planta* 250, 187–198. doi: 10.1007/s00425-019-03164-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Li, Zhang, Zhang, Li, Xu, Tian, Song, Luo, Xie, Wang, He, Xia, Zhang and Cao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Genome-Wide Association Study Pinpoints Quantitative Trait Genes for Plant Height, Heading Date, Grain Quality, and Yield in Rye (*Secale cereale* L.)

Dörthe Siekmann^{1,2}, Gisela Jansen³, Anne Zaar³, Andrzej Kilian⁴,
Franz Joachim Fromme² and Bernd Hackauf^{1*}

¹ Julius Kühn Institute, Federal Research Centre for Cultivated Plants, Institute for Breeding Research on Agricultural Crops, Sanitz, Germany, ² HYBRO Saatzeit GmbH & Co. KG, Schenkenberg, Germany, ³ Julius Kühn Institute, Federal Research Centre for Cultivated Plants, Institute for Resistance Research and Stress Tolerance, Sanitz, Germany, ⁴ Diversity Arrays Technology, Bruce, ACT, Australia

OPEN ACCESS

Edited by:

Shuyu Liu,
The Texas A&M University System,
United States

Reviewed by:

Shuhao Yu,
The Texas A&M AgriLife Research,
United States
Frank Maulana,
Louisiana State University Agricultural
Center, United States

*Correspondence:

Bernd Hackauf
bernd.hackauf@julius-kuehn.de

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 31 May 2021

Accepted: 22 September 2021

Published: 29 October 2021

Citation:

Siekmann D, Jansen G, Zaar A,
Kilian A, Fromme FJ and Hackauf B
(2021) A Genome-Wide Association
Study Pinpoints Quantitative Trait
Genes for Plant Height, Heading Date,
Grain Quality, and Yield in Rye (*Secale
cereale* L.).
Front. Plant Sci. 12:718081.
doi: 10.3389/fpls.2021.718081

Rye is the only cross-pollinating Triticeae crop species. Knowledge of rye genes controlling complex-inherited traits is scarce, which, currently, largely disables the genomics assisted introgression of untapped genetic variation from self-incompatible germplasm collections in elite inbred lines for hybrid breeding. We report on the first genome-wide association study (GWAS) in rye based on the phenotypic evaluation of 526 experimental hybrids for plant height, heading date, grain quality, and yield in 2 years and up to 19 environments. We established a cross-validated NIRS calibration model as a fast, effective, and robust analytical method to determine grain quality parameters. We observed phenotypic plasticity in plant height and tiller number as a resource use strategy of rye under drought and identified increased grain arabinoxylan content as a striking phenotype in osmotically stressed rye. We used DArTseqTM as a genotyping-by-sequencing technology to reduce the complexity of the rye genome. We established a novel high-density genetic linkage map that describes the position of almost 19k markers and that allowed us to estimate a low genome-wide LD based on the assessed genetic diversity in elite germplasm. We analyzed the relationship between plant height, heading date, agronomic, as well as grain quality traits, and genotype based on 20k novel single-nucleotide polymorphism markers. In addition, we integrated the DArTseqTM markers in the recently established ‘Lo7’ reference genome assembly. We identified cross-validated SNPs in ‘Lo7’ protein-coding genes associated with all traits studied. These include associations of the WUSCHEL-related homeobox transcription factor *DWT1* and grain yield, the DELLA protein gene *SLR1* and heading date, the Ethylene overproducer 1-like protein gene *ETOL1* and thousand-grain weight, protein and starch content, as well as the Lectin receptor kinase *SIT2* and plant height. A Leucine-rich repeat receptor protein kinase and a Xyloglucan alpha-1,6-xylosyltransferase count among the cross-validated genes associated with water-extractable arabinoxylan content. This study demonstrates the power of GWAS, hybrid breeding, and the reference genome

sequence in rye genetics research to dissect and identify the function of genes shaping genetic diversity in agronomic and grain quality traits of rye. The described links between genetic causes and phenotypic variation will accelerate genomics-enabled rye improvement.

Keywords: phenotyping, drought stress, DNA profiling, SNP, tiller number, SMART breeding, WOX transcription factor, arabinoxylan (AX)

INTRODUCTION

Rye (*Secale cereale* L.) is the only allogamous crop in the Triticeae tribe of the grasses. Natural genetic diversity in outbreeding rye enabled to achieve a series of technological advances that ultimately facilitated the establishment of hybrid breeding (Hackauf et al., 2021), a key technology for increasing and securing cereal production on finite arable land without increasing water and fertilizer use (Whitford et al., 2013). Hybrid breeding resulted in a strong response to the selection of favorable alleles for grain and quality traits in rye (Laidig et al., 2017) and contributed to keeping this orphan crop competitive in modern agricultural production systems. ‘Petkus’ and ‘Carsten’ represent two major germplasm pools exploited in hybrid rye breeding (Geiger and Miedaner, 2009). Beyond heterotic groups and just like in maize (White et al., 2020), the structure of commercial hybrid rye breeding is characterized by the largely isolated and unique sub-heterotic patterns of major breeding programs (Bauer et al., 2017; Vendelbo et al., 2020). As elite germplasm utilization across programs is impossible, capture and management of genetic diversity from germplasm resources are of outstanding importance for the long-term success of commercial hybrid rye breeding programs. However, the targeted identification of novel genetic variation in agronomically important genes, especially superior alleles occurring with low frequencies in self-incompatible rye germplasm collections, is currently largely disabled. This accounts for the fact that comparable progress in gene discovery like in barley (Pasam et al., 2012; Hill et al., 2021; Li M. et al., 2021) or wheat (Liu et al., 2017; Gao et al., 2021) has yet not been achieved in rye. With the recent release of two high-quality genome assemblies (Li G. et al., 2021; Rabanus-Wallace et al., 2021), rye has finally reached the genome era, enabling the integration and advancement of fundamental and applied breeding and research to understand how the genome builds, maintains, and operates rye. In order to accelerate the transition from merely phenotypic to haplotype-based breeding (Bevan et al., 2017; Brinton et al., 2020) and to substantially increase the efficiency, precision, and flexibility of rye breeding, further progress in rye genomic research is necessary to associate genome sequence information with phenotypes related to rye growth and development. This is particularly relevant to grain quality parameters, as the versatile uses of rye in the production of bread or mixed animal feeds have, so far, been considered to require highly divergent breeding goals (Kobylyansky et al., 2019). An efficient selection of grain quality parameters, particularly with respect to the content of arabinoxylans as the predominant dietary fiber in the rye grain, is currently a limiting factor in rye breeding.

Grain quality, as well as agronomic important traits controlling plant height, heading date, thousand-grain weight, or yield, reveal a continuous phenotypic variation and are genetically controlled by a network of multiple and interacting loci (Mackay et al., 2009). In rice, cloning of these quantitative trait loci (QTL) for grain yield components and other agronomic important traits (Yonemaru et al., 2010; Yamamoto et al., 2012; Li et al., 2018) has a significant impact on the genetic improvement of this important staple food (Xing and Zhang, 2010; Wang and Li, 2019). In contrast, comprehensive QTL analysis in interpool rye hybrids with a high heterosis level currently refers to a single biparental mapping population from the ‘Carsten’ gene pool crossed to a CMS tester from the ‘Petkus’ gene pool (Hackauf et al., 2017a; Miedaner et al., 2018). Compared to QTL mapping with biparental populations, genome-wide association studies (GWAS) offer a sampling of greater levels of genetic diversity and higher resolution of QTL positions using already existing breeding lines and genetic stocks (Jamann et al., 2015). In rye, GWAS have been reported for the traits leaf rust resistance, pre-harvest sprouting resistance, and α -amylase activity (Rakoczy-Trojanowska et al., 2017), Tan spot resistance (Sidhu et al., 2019), Fusarium head blight resistance (Gaikpa et al., 2020), and stem rust resistance (Gruner et al., 2021).

Here, we describe the first comprehensive GWAS for plant architecture, grain quality, and yield in rye. We aimed to (i) assess the genetic variation for plant height, heading date, tiller number, grain yield, grain weight, grain protein, and starch, as well as total and water-extractable arabinoxylan content in rye, (ii) identify QTL for these traits by GWA mapping and estimate their effects, (iii) investigate the co-localization between QTL for grain yield and yield parameters, as well as between grain quality parameters. Furthermore, we used the recently released ‘Lo7’ genome assembly (Rabanus-Wallace et al., 2021) to describe the genetic architecture of the assessed traits.

MATERIALS AND METHODS

Plant Materials

The genetic materials analyzed in this study encompass two data sets of advanced-cycle inbred lines of a commercial hybrid rye breeding program. The two data sets represented a gametic sample of both heterotic gene pools and comprised a total of 126 S5-lines of the ‘Petkus’ and 15 restorer synthetics of the ‘Carsten’ gene pool. Male-sterile BC2-analogs of S5-lines from the ‘Petkus’ pool have been developed by backcrossing into the CMS-inducing Pampa (P) cytoplasm (Geiger and Schnell, 1970). We have used the line x tester mating design of Kempthorne

(1956) to develop two large multiple-hybrid populations using the male-sterile BC2-analogs as females to be tested and the restorer synthetics as testers. A total of 298 top cross hybrids were obtained in the first data set by pollinating 79 BC2-analogs from the 'Petkus' pool with 12 synthetics from the 'Carsten' pool. In the second data set, 305 top cross hybrids originated from 84 BC2-analogs from the 'Petkus' pool pollinated by 11 synthetics from the 'Carsten' pool. Both data sets were connected by 77 common experimental hybrids. These plant materials are proprietary to HYBRO Saatzzucht GmbH & Co. KG. The released hybrid varieties 'Rasant', 'Askari', 'Minello', 'Brasetto', 'Palazzo', 'Visello', as well as 'SU Mephisto', 'Palazzo', 'Visello', 'Brasetto', 'SU Allawi', and 'Minello', were used as checks in 2011 and 2012, respectively.

Phenotypic Data Analyses

Experimental hybrids of data set 1 were evaluated in 2011 at 10 locations, representing target environments of rye cultivation in Germany and Poland. Soils with different qualities in agricultural terms enabled to define two locations each at the sites Kleptow and Wulfode as measured on a scale of soil value of arable land (German: Bodenwertzahl), which is determined from soil sampling data and that ranges from 0 (very low) to 100 (very high). The soil type at Kleptow is a loam, partly with loess covering (soil value 60), while Kleptow-Sand is a heavy-to-clayey loam (soil value 45), just like at Wulfode (soil value 38). The soil at Wulfode-Sand is a sandy loam (soil value 25). Thus, the locations comprise (1) Kleptow (KLE) N53.4°, E13.9°, (2) Kleptow-Sand (KLS) N53.4°, E13.9°, (3) Wulfode (WUL) N53°, E10.2°, (4) Wulfode-Sand (WLS) N53°, E10.2°, (5) Cappeln (WES) N52.8°, E8°, (6) Bornhof (BOH) N53.5°, E12.9°, (7) Granskevitz (GKV) N54.5°, E13.2°, (8) Sulejów (SUL) N51.4°, E19.8°, (9) Poznań (POZ) N52.4°, E16.6°, and (10) Uhnin (UHN) N51.6°, E23°. In 2012, the location WLS was not included. The location $\times$ year combinations are subsequently referred to as environments. The year 2011 and, in particular, the period from January to May was characterized by natural drought stress in Europe with many areas receiving <40% of long-term mean annual precipitation [DWD (Deutscher Wetterdienst), 2012]. In contrast, no precipitation anomalies in relation to the long-term mean for 1951–2000 have been recorded for the hydrological period from November 2011 until May 2012 in the North German Plain and Poland [DWD (Deutscher Wetterdienst), 2012]. Entries were allocated to trials laid out as α -lattice designs with two replicates on 5 to 6 m² plots, connected by elite hybrid checks. Test cross performance was evaluated for grain yield (GYD, Mg ha⁻¹), plant height (PHT, cm), thousand-grain weight (TGW, g), heading date (HDT, 1 = late – 9 = early), tiller number (TIN), grain protein (GPC, g/kg), and starch (STC, g/kg), as well as total arabinoxylan (TAX, g/kg), and water-extractable arabinoxylan (WAX, g/kg) content. For TGW data across 10 environments (KLE2011, KLS2011, BOH2011, WLS2011, POZ2011, KLE2012, KLS2012, BOH2012, WUL2012, and WES2012) were recorded. GPC, STC, TAX, and WAX were assessed in eight environments (KLE2011, KLS2011, BOH2011, WLS2011, POZ2011, KLS2012, KLE2012, and WUL2012), while TIN was assessed in three environments (KLE2011, KLS2011,

and KLE2012). Grain yield was adjusted to a moisture content of 140 g H₂O kg⁻¹. We have used a near-infrared spectroscopy (NIRS) calibration (Agelet and Hurburgh, 2010) to predict grain quality parameters in experimental rye hybrids. This NIRS calibration was adjusted in both data sets by grain samples selected from the experimental hybrids of both years. Subsequent to non-destructive NIRS scans, the content of GPC, TAX, as well as WAX, was assayed as recently described (Jürgens et al., 2012). Native starch content was determined according to the Ewers polarimetric method (ISO 10520) using a Polartronic universal polarimeter (Schmidt + Haensch, Berlin, Germany).

For each data set, best linear unbiased estimators (BLUEs) of the experimental hybrids were calculated over locations according to model Equation (1):

$$y_{ijkl} = \mu + g_i + \alpha_j + \delta_{ij} + \beta_{jk} + \varphi_{jkl} + \varepsilon_{ijkl} \quad (1)$$

where y_{ijkl} is the phenotypic observation of the i th hybrid at the j th location in the l th in complete block of the k th replication; μ , the intercept; g_i , the genetic effect of the i th hybrid; α_j , the effect of the j th location; δ_{ij} , the genotype-by-environment interaction; β_{jk} , the effect of the k th replicate at the j th location; φ_{jkl} , the effect of the l th incomplete block of the k th replication at the j th location; and ε_{ijkl} , the residual. μ and g_i were regarded as fixed, all other effects as random. For estimation of variance components, g_i was defined as random. Broad-sense (entry-mean) heritability (H^2) was estimated based on Equation (2):

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{G \times E}^2}{E} + \frac{\sigma_\varepsilon^2}{ER}} \quad (2)$$

with σ_G^2 , $\sigma_{G \times E}^2$, and σ_ε^2 being the variances of genotype, genotype-by-environment interaction, and residual plot error, respectively. E denotes the number of environments and R the number of replications per environment. Analysis of phenotypic data was performed using package lme4 in R (R Core Team, 2017). Estimation of variance components was conducted using the REML procedure in R. Pearson's correlation coefficients between phenotypic traits were calculated based on BLUEs with R package metan and visualized using package pheatmap in R.

Genotyping and Linkage Map Construction

DNA profiles were established for a panel of 180 'Petkus' and 34 'Carsten' genotypes as well as 91 plants of the RIL-population L2039-N $\times$ DH (Martis et al., 2013; Hackauf et al., 2017a). DNA samples were genotyped using DArTTM and DArTseqTM-technology as previously described (Bolibok-Bragoszewska et al., 2009; Targońska-Karasek et al., 2017) at Diversity Arrays Technology Pty. Ltd., Yarralumla ACT, Australia. Marker genotypes of the single cross hybrids were inferred from the parental genotypes. For genetic map construction, markers with more than 10% missing values and more than 5% heterozygote genotypes were excluded from the analysis. Perfectly identical marker loci were identified in JoinMap 4 (Van Ooijen, 2006) and removed from the core data set to reduce calculation efforts. Markers with a maximum of 1% missing values were used to establish a framework map, defining the number, joint genotype,

and position of genomic bins. Grouping of DArTseq™ markers, together with DArT™, single-nucleotide polymorphism (SNP), and SSR markers, from the L2039-N x DH map (Martis et al., 2013) was conducted using QTL IciMapping 3.2 (Meng et al., 2015). The high-density linkage map was constructed using the locus-ordering algorithm RECORD (Van Os et al., 2005). Map quality was evaluated by computing the sum of adjacent recombination fractions (SARF) with a window size of 5. DArTseq™ SNP tags were integrated into the ‘Lo7’ genome assembly (Rabanus-Wallace et al., 2021) by using the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990). A significance threshold with more than 98% identity of at least 65 bp was used. Cloned and functionally characterized rice genes (Yonemaru et al., 2010) were queried by BLASTP against the ‘Lo7’ peptide sequences (Rabanus-Wallace et al., 2021) with an E-value < 1e-20 as a significance threshold. BLASTN sequence similarity searches were conducted as described (Hackauf et al., 2009) to integrate first-generation DArT™ as well as EST-derived markers flanking previously identified QTL (Hackauf et al., 2017a; Miedaner et al., 2018) in the ‘Lo7’ genome assembly.

Analysis of Genetic Diversity, Linkage Disequilibrium, Kinship, Principal Coordinates, and Population Structure

For the molecular characterization of elite rye inbred lines, SNPs with a call rate > 0.8 and a minor allele frequency (MAF) > 0.05 were selected using the R package snpReady. The basic population genetic statistics observed heterozygosity (H_{obs}), gene diversity (expected heterozygosity, $\hat{H}$, Nei, 1987), inbreeding coefficient (F_{IS}), as well as the Analysis of Molecular Variance (AMOVA, Excoffier et al., 1992), were calculated using the R package hierfstat. A Cavalli-Sforza and Edwards (1967) chord distance matrix was calculated with the R package hierfstat and analyzed in a Principal Coordinates Analysis (PCoA) as well as hierarchical clustering using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) algorithm using the R packages ape and stats, respectively. The genetic differentiation between both heterotic gene pools was inferred based on the pairwise F_{ST} estimator proposed by (Weir and Cockerham, 1984). For evaluation of linkage disequilibrium (LD) in the two phenotypic data sets, mapped SNP markers with MAF > 5% were used to calculate squared allele frequency correlation (r_{vs}^2) values with the R package LDcorSV. To correct the estimation of LD for relatedness between the single hybrids, kinship coefficients were calculated with dominant DArT™ and SilicoDArT™ markers according to Hardy (2003) using the software SPAGeDi (Hardy and Vekemans, 2002). To eliminate negative kinship values while keeping variation, all coefficients were adjusted for the lowest value instead of setting all negative coefficients to zero. For analysis of population structure, five independent runs were performed for $K = 1-9$ using 1,523 dominant DArT™ markers in data set 1 and 4,000 randomly chosen SilicoDArT™ markers in data set 2, with the software Structure v2.2.3 Pritchard et al. (2000), with a burn-in of 100,000 and 200,000 iterations. The probable number of populations was determined according to Evanno et al. (2005). Both factors, kinship and population

structure, were taken into account during the calculation of r_{vs}^2 values. Non-linear regression was used to analyze the decay of LD with genetic distance for each chromosome separately (Remington et al., 2001). A critical value for the determination of LD decay was calculated in R from the distribution of square root-transformed r_{vs}^2 values of the unlinked loci. The 95th percentile of that distribution served as the threshold value, beyond which LD was likely to be caused by genetic linkage. Visualization of LD calculation results was done with the R package LDheatmap. For PCoA in the two sets of experimental hybrids, polymorphic SilicoDArT™ markers with a call rate > 0.95 were used to calculate Cavalli-Sforza and Edwards (1967) chord distance matrices. All visualizations were done with the packages ggplot2 and ggtree in R.

Genome-Wide Association Mapping

Single-nucleotide polymorphism markers were filtered for MAF > 5% and a call rate > 0.2 in TASSEL 3 (Bradbury et al., 2007) to identify marker/trait associations in the two data sets. Taking the cofactors kinship and population structure into account, a mixed linear model approach according to model Equation (3) was used in combination with the P3D algorithm (“population parameters previously determined”) to determine significant associations in TASSEL:

$$y = X\beta + Zu + e \quad (3)$$

where y is the vector of observations; β represents an unknown vector containing fixed effects, including genetic marker and population structure; u is an unknown vector of random additive genetic effects from multiple-background QTL for individuals; X and Z are the known design matrices; and e is the unobserved vector of random residual. Vector u and e are assumed to be normally distributed with null mean and variance of $\text{Var}_e^u = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}$, where $G = \sigma_a^2 K$ with σ_a^2 as additive genetic variance and K as kinship matrix. R is the residual effect with $R = I\sigma_e^2$, where σ_e^2 is the residual variance.

To adjust raw p -values for multiple hypotheses testing, the false discovery rate was controlled according to Benjamini and Hochberg (1995) using the R package multtest. The localization of associated SNPs was illustrated using MapChart (Voorrips, 2002). Phenotypic variances (R^2) explained by individual SNPs were plotted using ggplot2 in R.

RESULTS

Elite Rye Germplasm Reveals Pronounced Genetic Variation in Agronomic and Quality Traits

Analysis of variance (ANOVA) revealed that genotypic effects were statistically significant ($p < 0.05$) for all traits except TIN in 2011 (Table 1). All traits but GPC, STC, WAX, and TAX in 2012 as well as TIN showed significant ($p < 0.05$) genotype x environment (GxE) interaction variances. The GxE interaction variances of GYD in both years, as well as of WAX in 2011 and HDT in 2012, were larger than the variances of the genotype

TABLE 1 | First and second order statistics of experimental rye hybrids.

Year	2011						2012					
	$\mu \pm s$	$x_{\min}-x_{\max}$	σ^2_G	$\sigma^2_{G \times E}$	σ^2_e	H^2	$\mu \pm s$	$x_{\min}-x_{\max}$	σ^2_G	$\sigma^2_{G \times E}$	σ^2_e	H^2
GYD [Mg/ha]	8.56 $\pm$ 0.28	7.82–9.4	0.05*	0.13*	0.18	0.69	9.31 $\pm$ 0.35	8.35–10.16	0.09*	0.23*	0.16	0.71
PHT [cm]	126.9 $\pm$ 4.9	114.1–138.4	22.03*	7.14*	13.92	0.94	137.5 $\pm$ 4.8	126.8–149.1	21.65*	5.23*	13.9	0.94
HDT [1–9]	6.03 $\pm$ 0.51	4.65–7.4	0.22*	0.18*	0.32	0.87	6.23 $\pm$ 0.4	4.69–7.14	0.11*	0.26*	0.3	0.72
TIN [spikes/m ²]	526.64 $\pm$ 35.6	423.5–630.2	339.4	0	2926.8	0.32	637.63 $\pm$ 60.5	446.8–806.8	898.3*	NA	3734.6	0.33
TGW [g]	39.1 $\pm$ 1.6	35.7–43.3	2.19*	0.87*	1.22	0.88	32.55 $\pm$ 1.96	27.3–38.6	3.47*	0.96*	1.42	0.91
GPC [g/kg]	89.4 $\pm$ 3.48	80.1–100.6	10.1*	2.04*	10.3	0.88	88.96 $\pm$ 3.1	79.6–98.2	6.2*	1.2	11.3	0.73
STC [g/kg]	576.7 $\pm$ 4.86	562.2–589.2	20.2*	5.07*	20.2	0.87	570.7 $\pm$ 7.3	546.9–589	31.1*	1.8	99.6	0.64
WAX [g/kg]	23.9 $\pm$ 1	20.9–27	0.73*	0.81*	0.79	0.75	16.3 $\pm$ 1.2	12.5–20	0.87*	0	2.7	0.66
TAX [g/kg]	109.1 $\pm$ 3.7	98–119.1	12.1 *	4.2*	6.8	0.89	87.1 $\pm$ 3.9	76.7–96.5	12.2*	1.8	12.2	0.82

Means and standard deviation ($\mu \pm s$), minimum and maximum ($x_{\min}-x_{\max}$), estimates of variance components (genotype, σ^2_G ; genotype $\times$ environment interaction, $\sigma^2_{G \times E}$; error, σ^2_e), and entry-mean heritabilities (H^2) for agronomic and quality traits evaluated across 3–19 environments in 2011 and 2012. GYD, grain yield; PHT, plant height; HDT, heading date; TIN, tiller number; TGW, thousand-grain weight; GPC, grain protein content; STC, starch content; TAX, total arabinoxylan content; WAX, water-extractable arabinoxylan content. * $p = 0.05$.

(Table 1). The average performance in grain yield (–8%), TIN (–17%), and PHT (–8%) was lower in 2011 as compared with 2012. In contrast, the average TGW (+20%), as well as the content of TAX (+25%) and WAX (+47%), was higher in 2011 as compared with 2012. High heritability estimates depict a low error variance and a high genetic variance for all traits with the exception of TIN ($H^2 = 0.3$). Q-Q plots are fairly straight and indicate that normal distribution is a reasonably good approximation for all traits (Figure 1).

The pattern of phenotypic correlation coefficients is illustrated in Figure 2. Most significant ($p < 0.001$) and consistent correlations were found between STC and PHT, STC and GPC, STC and TAX, as well as between TAX and WAX. The significant ($p < 0.01$) correlations between GYD and STC as well as GPC are in opposite directions in both years. GYD, TGW, and HDT, as well as TGW, STC, GPC, TIN, and PHT, revealed significant ($p < 0.001$) correlations in 2012 only. Similarly, the correlation between STC and WAX was significant ($p < 0.001$) in 2012 only. Plant height was significantly ($p < 0.001$) correlated with TGW in 2012 and starch content in both years but neither correlated with GYD in 2011 nor in 2012. A very strong negative correlation ($r = -0.9$) could be observed between GPC and STC in 2011 only. No moderate correlations were found except the correlations previously mentioned (Figure 2).

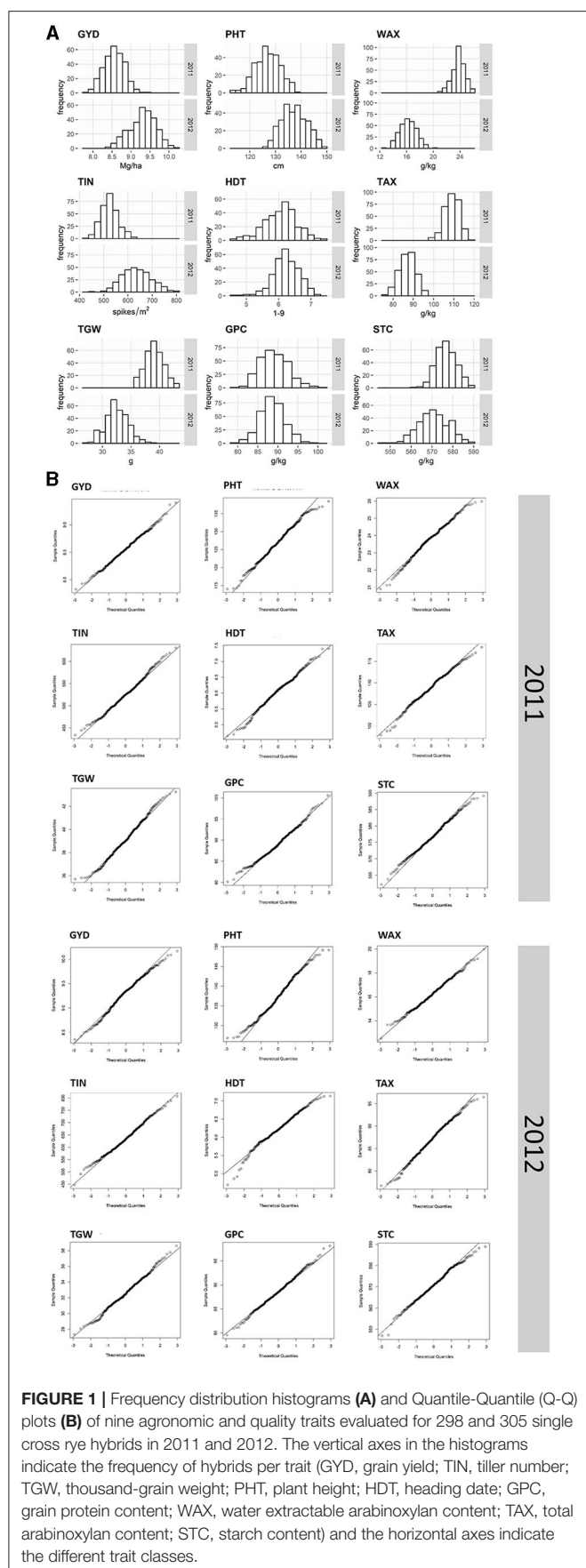
Significant Molecular Differentiation of Elite Rye Germplasm

A total of 2,965 SNPs were successfully called with high quality in 214 seed and pollen parent lines. In total, 2,006 of these SNPs (67.7%) could be mapped in the ‘Lo7’ reference genome sequence and revealed an equal distribution on the seven ‘Lo7’ pseudomolecules. The average PIC of these SNPs was 0.258, ranging from 0.09 to 0.38. A detailed list of these informative SNP loci, including physical map position in the ‘Lo7’ genome assembly, base change, MAF, heterozygosity, gene diversity, and PIC, is provided in Supplementary Table 1. The average heterozygosity of each ‘Petkus’ line was 6%, while, for the ‘Carsten’ genotypes, an average heterozygosity of 29% (Table 2)

was determined. Analysis of Molecular Variance (AMOVA) results revealed that molecular variation was mainly (57.1%) found among individuals within populations as expected for cross-pollinated species, whereas variation observed among populations within groups explained 18.5% of the total genetic variability (Table 3). The calculated fixation index $F_{ST} = 0.185$ of Weir and Cockerham (1984) emphasized a significant differentiation between both heterotic populations. Cavalli-Sforza and Edwards (1967) chord genetic distances between pairwise comparisons of all the 214 lines ranged from 0.012 to 0.487, and the overall average distance was 0.356. The UPGMA tree clustered this population into two major groups, perfectly in line with the well-known heterotic groups in rye (Figure 3). The major group was composed solely of ‘Petkus’ lines and was divided into five clades comprising 18–84 lines each. In the second monophyletic branch, all ‘Carsten’ lines are grouped (Figure 3). PCoA on the entire set of 214 lines (Figure 4) showed a clear separation of the two groups identified in the cluster analysis. The first two principal coordinates (PCos) from PCoA explained 22.3% of the total SNP variation among the samples.

Family Substructure and Rapid Decay of LD in Experimental Rye Hybrids

Based on Evanno’s DeltaK method, we determined $k = 4$ as the most probable number of groups in each of the two sets of experimental hybrids (Figure 5). The principal coordinates analysis confirmed this major population substructure and revealed that the first two principal coordinates explained 42.3 and 42.8% of the genetic variation within the 298 and 305 rye hybrids (Figure 6). The grouping of experimental hybrids into subpopulations referred mainly to the individual pollen parent genotypes. We have integrated 13,337 SilicoDARTTM and 3,711 novel SNP markers in the genetic linkage map for the RIL-population L2039-N $\times$ DH. The advanced map covered 1946.4 cM of the rye genome with map length of the seven rye chromosomes varying between 227 cM of chromosome 3R and 340 cM of chromosome 4R (Supplementary Table 2). The 3,711 novel SNP markers were equally distributed across the genome



with SNP numbers varying between 402 on chromosome 3R to 577 on chromosome 4R (**Supplementary Table 2**). We have used this high-density map to investigate the LD in the analyzed experimental hybrids. Visual inspection of intra-chromosomal heatmaps from all segregating markers depicts no noticeable differences in the LD structure between both germplasm collections (**Figure 7**). In the distal region of chromosome 6RS, a small distinct LD block was observed in both populations. Mean r_{vs}^2 of individual chromosomes ranged from 0.07 for 1R in 2011 to 0.087 for 4R in 2011 (**Table 4**). In the collection of 298 single cross hybrids 2011, the critical value of r_{vs}^2 (basal LD) is $r_{vs}^2 = 0.2467$. Similarly, in the cohort of 305 single cross hybrids 2012, the critical value is $r_{vs}^2 = 0.2208$. Based on these critical values, the intrachromosomal LD decayed between 1.51 (2R in 2011) and 4.6 cM (6R in 2012) for the individual chromosomes in the germplasm collection (**Table 4**). The mean LD decay over the whole genome was calculated as 2.2 cM (2011) and 2.77 cM (2012).

GWAS of Agronomic and Quality Traits

Out of 20,232 DArTseqTM-derived SNP tags, 228 (1.1%) are covered by the Rye 600k genotyping array (**Supplementary Table 3**) and 10,712 (52.9%) mapped to the 'Lo7' pseudomolecules (**Supplementary Table 2**). Map positions could be determined for 10,171 (94.9%) of these SNP tags with 1,302 located on chromosome 1R, 1,534 on 2R, 1,388 on 3R, 1,498 on 4R, 1,474 on 5R, 1,464 on 6R, and 1,520 on 7R (**Supplementary Table 2**). About 3,901 SNPs (36.4%) represent predicted coding sequences in the 'Lo7' genome assembly. Comparison between 2,149 (21.1%) genetically mapped SNPs revealed almost perfect collinearity to the physical map. GWAS revealed that between 25 of these SNPs on chromosome 1R and 278 on chromosome 6R were significantly ($p_{adj} < 0.05$) associated with PHT, HDT, and the studied agronomic traits (**Supplementary Table 4**). We identified between 34 on chromosome 1R and 186 on chromosome 7R significantly ($p_{adj} < 0.05$) associated SNPs for quality traits (**Supplementary Table 4**). For GYD, 38 SNP loci showed significant association ($p_{adj} < 0.05$) in both years. For all remaining traits, this value ranged between 15 and 132 significant SNPs (**Table 5**, **Figure 8**, and **Supplementary Table 4**). The explained phenotypic variance (R^2) of the individual QTL ranged from 1.4 to 29.9% for GYD, 1.3 to 31.3% for PHT, 1.4 to 31.6% for HDT, 1.4 to 26.3% for TIN, 1.4 to 42.7% for TGW, 1.4 to 27.3% for GPC, 1.4 to 24.7% for STC, 1.3 to 38.8% for TAX, and 1.4 to 30.1% for WAX (**Figure 9**, **Supplementary Table 4**). In total, between 100 and 566 of the associated SNPs represent predicted protein-coding genes of the 'Lo7' assembly (**Supplementary Table 4**). Between 9 and 35 marker-trait associations (MTA) in protein-coding sequences could be cross-validated in both datasets (**Table 5**, **Figure 8**). Nine of the cross-validated MTA for PHT are located within the *QPh3-7R* interval, including SECCE7Rv1G0469090, SECCE7Rv1G0471520, and SECCE7Rv1G0472410, respectively. SECCE6Rv1G0401900 ranks among two MTA for HDT located within *QHdt-6R*. Among the MTA for TGW, we identified the polyubiquitin-encoding gene SECCE5Rv1G0370510 residing within *QTgw-5R*. SECCE6Rv1G0382730, encoding an

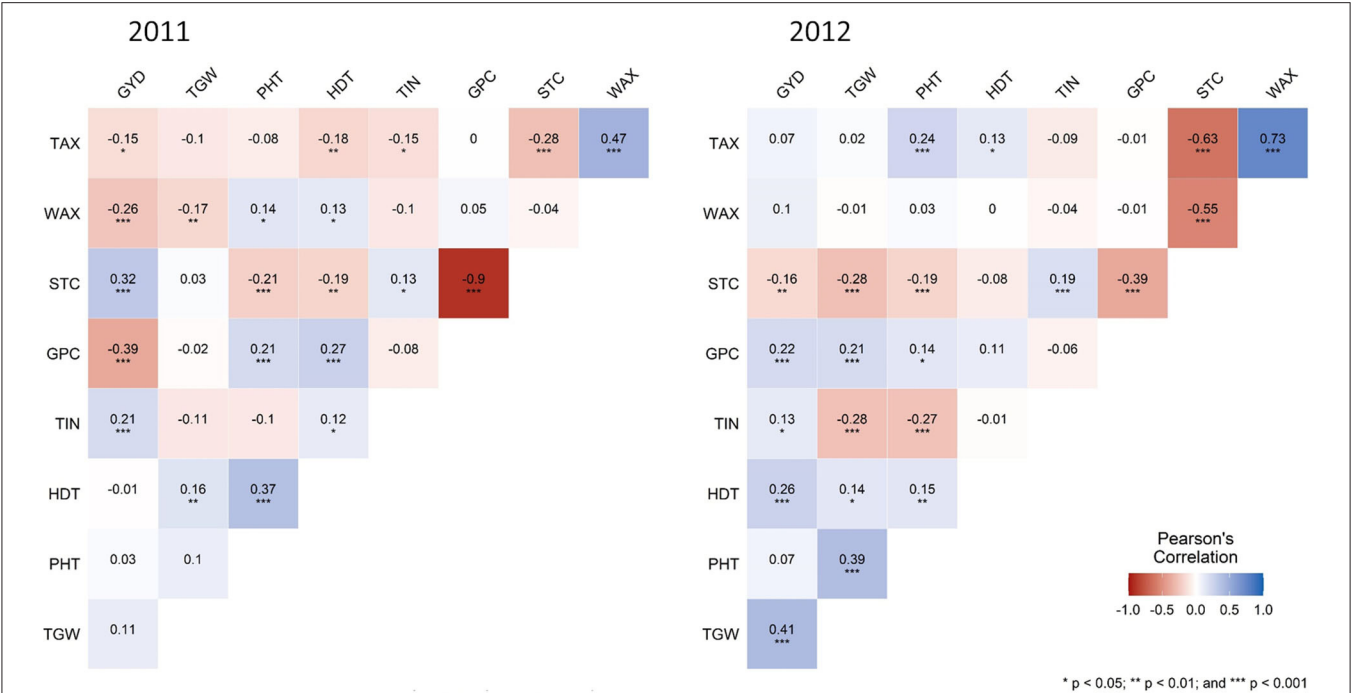


FIGURE 2 | Heatmaps of Pearson's correlation coefficients for nine agronomic and quality traits in rye. Blue color indicates positive correlation and red color indicates negative correlation among different traits, with color intensity variance depicting the strength of correlation. GYD, grain yield; TIN, tiller number; TGW, thousand-grain weight; PHT, plant height; HDT, heading date; GPC, grain protein content; WAX, water extractable arabinoxylan content; TAX, total arabinoxylan content; STC, starch content. Categorization: $0.25 \leq r < 0.45$ weak, $0.45 \leq r < 0.65$ moderate, $0.65 \leq r < 0.85$ strong, $0.85 \leq r$ very strong. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, blank for non-significant.

TABLE 2 | Genetic diversity of elite rye inbred lines representing the heterotic gene pools 'Petkus' and 'Carsten'.

Population name	Breeding level	Individuals	A_p	H_{obs}	H	F_{IS}
'Petkus'	Elite seed parent	180	247	0.06	0.30	0.77
'Carsten'	Elite pollen parent	34	26	0.29	0.31	0.07

The level of genetic diversity of each population was estimated based on 2,965 SNP markers and described with the parameters number of private alleles (A_p), observed heterozygosity (H_{obs}), gene diversity (expected heterozygosity, H), and inbreeding coefficient (F_{IS}).

Embryo-defective protein, mapped to *QTgw-6R*. We illustrate the identification of putative causal genes associated with variation in target traits of rye breeding using plant height, thousand-grain weight, and yield, as well as arabinoxylan content as examples.

Candidate Genes for Plant Height

We observed 483 SNPs associated with PHT in 'Lo7' protein-coding genes (Supplementary Table 4). Fifteen (3.1%) of these genes are orthologs of cloned rice QTL, including the Gibberellin-insensitive dwarfing gene *SLR1*, the epigenetic regulator *Decrease in DNA Methylation 1* (*DDM1*), the lectin receptor-like kinase *SIT2*, the phosphate transporter *PT8*, and the transmembrane ABC transporter

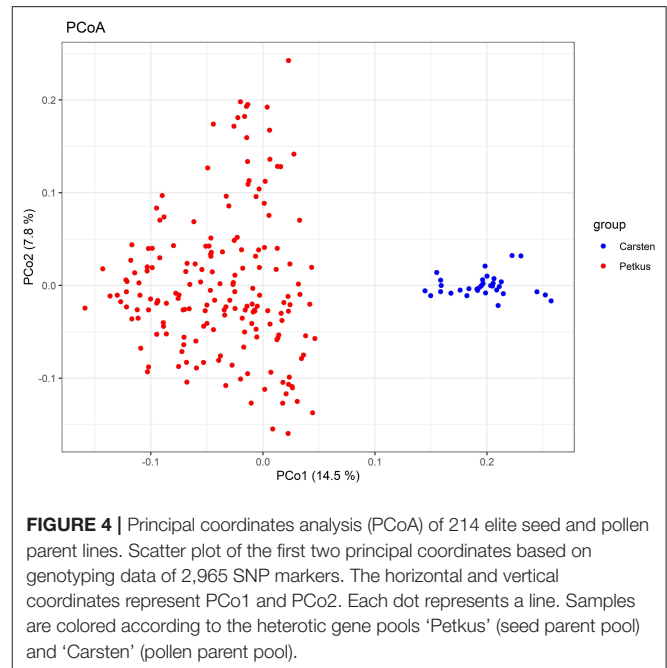
TABLE 3 | Analysis of molecular variance (AMOVA) for the total breeding population.

Source of variation	Variance component	Proportion of explained variation
Among populations	203.26	0.185
Within populations	627.30	0.571
Within individuals	268.33	0.244
Total	1098.89	

STAR2 (Supplementary Table 5). *SIT2* turned out to be associated in both years. Altogether, 35 SNPs in 'Lo7' gene models were associated in both years (Table 5). The SNP in SECCE1Rv1G0063100 encoding a putative translation initiation factor IF-2 explains more than 10% of the phenotypic variance in both years.

Candidate Genes for Thousand-Grain Weight

We detected 685 SNPs associated with TGW in 'Lo7' protein-coding genes (Supplementary Table 4). Fourteen (2%) of the 'Lo7' protein-coding genes are orthologs of cloned rice QTL, including the zinc finger protein gene *ZFP179*, a cytochrome P450 protein-encoding gene, and



an *ETHYLENE OVERPRODUCER 1-like* gene (*OsETOL1*) (Supplementary Table 5). The *OsETOL1* ortholog of SECCE7Rv1G0479810 affects TIN, GPC, and STC in rye as well. Contrasting effects on TIN and STC as compared with TGW and GPC could be observed for the A547G SNP in SECCE7Rv1G0479810 (Figure 10). The experimental hybrids evaluated in 2012 represented three SNP genotypes, while, in 2011, only two genotype classes could be detected. Major effects ($R^2 > 10\%$) on TGW could be observed for 9 of the 525 SNPs (1.7%) associated in 2011, including the predicted sucrose transporter SECCE7Rv1G0456720 and for 535 of the 1,465 (36.5%) SNPs associated in 2012 (Supplementary Table 4). The cross-validated effects of 132 SNPs on thousand-grain weight explained between 1.4 and 30.3% of the phenotypic variance. The protein-coding sequences represented by these SNPs include SECCE1Rv1G0038830, a nitrate/peptide transporter; SECCE1Rv1G0052830, encoding a methyltransferase, the invertase inhibitor SECCE3Rv1G0153090; SECCE5Rv1G0376740, encoding a basic helix-loop-helix (bHLH) DNA-binding superfamily protein; SECCE6Rv1G0427440, encoding a jasmonate O-methyltransferase; and SECCE7Rv1G0479810, encoding an ethylene-overproduction protein.

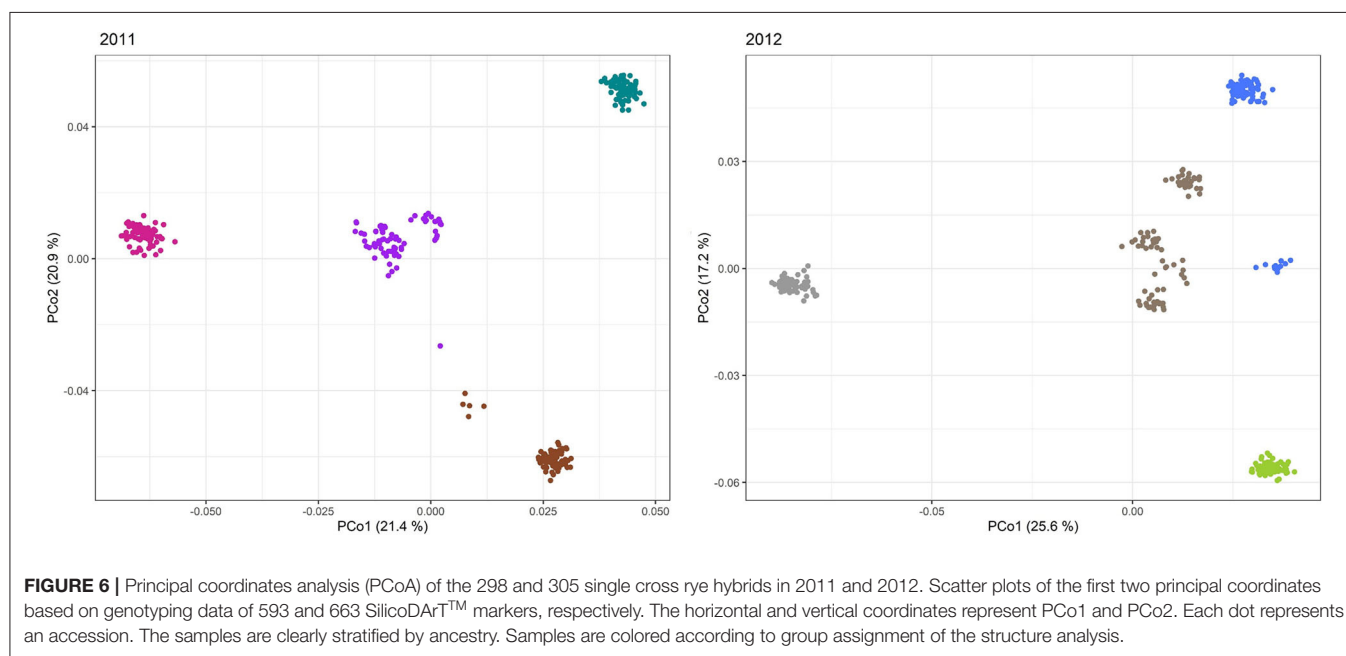
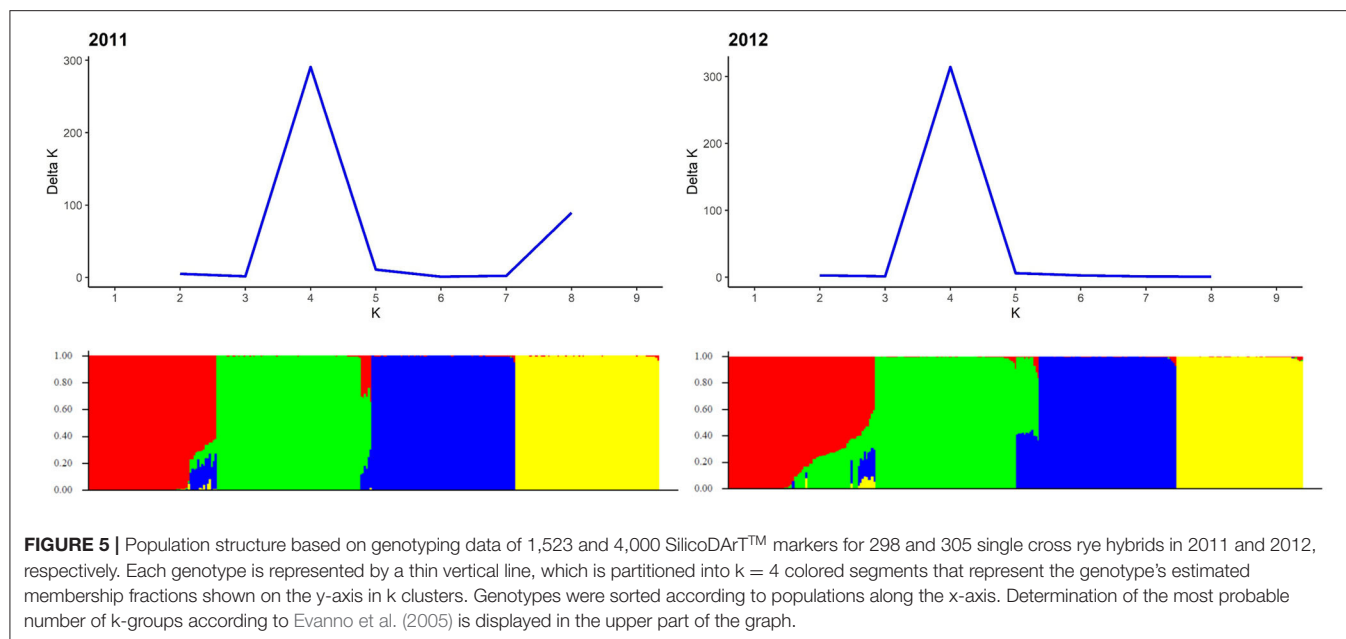
Candidate Genes for Grain Yield

We identified 19 (3.8%) out of 500 SNPs associated with GYD in protein-coding genes of the 'Lo7' assembly as orthologs of cloned rice QTL (Supplementary Table 5). These included the WUSCHEL—related homeobox transcription factor *DWT1*, the Zinc-finger transcription factor *YABBY1*, the gene *DTH2* promoting heading under long-day conditions, the cellulose synthase *bc6*, the soluble starch synthase IIa, the Cytochrome

P450CYP96B4, the metal-nicotianamine transporter *YSL2*, the amino acid transporter AAP6, the ATP-binding cassette protein, encoding gene *ABCB14*, the putative potassium efflux antiporter *Albino midrib 1* (*AM1*), the Zinc finger family protein encoding gene *sor1*, and the transmembrane ABC transporter *STAR2*. While the SNP representing the *DWT1*-ortholog in rye was associated in both years, the SNPs in the rye orthologs of *YABBY1* and *ABCB14* were associated in 2011 and the remaining SNPs in 2012. Among the 337 SNPs associated in 2011, eight (2.4%) explained more than 10% of the phenotypic variance. Three of these SNPs represent genes encoding a cytochrome P450 protein, a putative late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein, as well as an ARM repeat superfamily protein. Likewise, 101 (20.2%) among the 500 SNPs in protein-coding sequences associated in 2012 had major effects ($R^2 > 10\%$) on GYD. Except for 6, the 38 SNPs associated in both years had minor effects ($R^2 < 10\%$) on grain yield. The genes represented by these major effect SNPs include SECCE3Rv1G0183650 encoding, a WUSCHEL homeobox protein, SECCE5Rv1G0323860.1, a HAESA-like LRR receptor protein kinase, SECCE6Rv1G0412350, a flavonol synthase, SECCE6Rv1G0416090, a putative Snf1-related kinase interactor, SECCE7Rv1G0469060, a Gibberellin 2-beta-dioxygenase, and SECCEUnv1G0531960, a BEL1-like homeobox protein.

Candidate Genes for Arabinoxylan Content

We discovered 416 and 433 SNPs associated with TAX and WAX, respectively, in 'Lo7' protein-coding genes (Supplementary Table 4). Eleven (2.6%) and nine (2.1%) of the 'Lo7' protein-coding genes are orthologs of cloned rice QTL, including the cZ-O-glucosyltransferase *cZOGT1*, an embryo and endosperm development-involved



cyclin and the transmembrane ABC transporter *STAR2* (**Supplementary Table 5**). Major effects ($R^2 > 10\%$) on TAX and WAX in 2011 were observed for 17 and 23 SNPs, respectively. Corresponding protein-coding sequences represented, among others, the receptor protein kinases SECCE2Rv1G0135820 and SECCE4Rv1G0223400, and the hexosyltransferase SECCE5Rv1G0328330. In 2012, 249 and 202 SNPs revealed major effects ($R^2 > 10\%$) on TAX and WAX. These SNPs described, among others, the α -glucosidase SECCE3Rv1G0149680 and the xyloglucan α -1,6-xylosyltransferase SECCE6Rv1G0407620.

Major effects ($R^2 > 10\%$) on TAX were observed for SNPs in the Glycosyltransferase-encoding genes SECCE3Rv1G0197970 and SECCE7Rv1G0456850, and minor effects ($R^2 < 10\%$) on TAX and WAX for SNPs in six further Glycosyltransferase-encoding genes (**Supplementary Table 4**). The predicted glycosyltransferase SECCE3Rv1G0150570 revealed a significant association with TAX and WAX content. Across eight environments, we detected 31 and 15 cross-validated SNPs in protein-coding genes associated with TAX and WAX, respectively (**Table 5**, **Supplementary Table 4**).

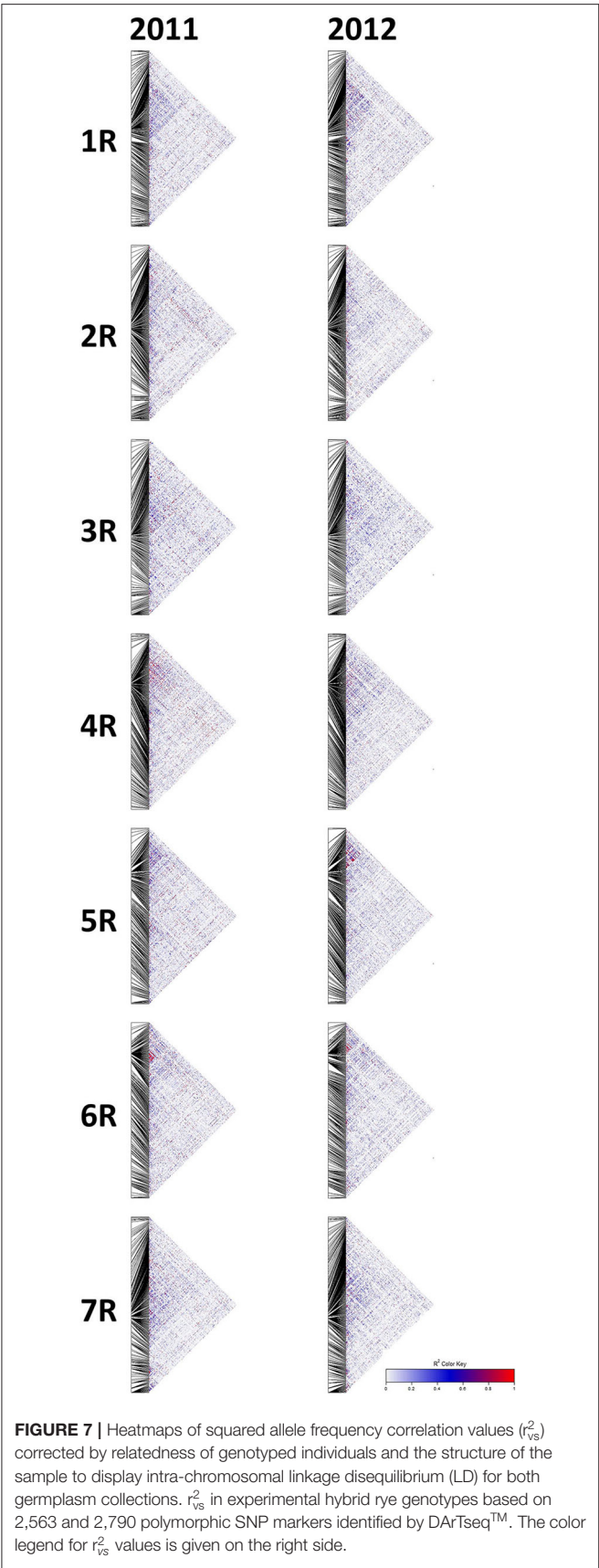


TABLE 4 | Estimates of LD as well as distances at which LD decays below background LD for single cross hybrids in 2011 and 2012.

Year	2011		2012	
Chromosome	Mean r^2_{vs}	LD decay	Mean r^2_{vs}	LD decay
1R	0.070	1.75	0.082	3.55
2R	0.077	1.51	0.072	1.81
3R	0.079	1.7	0.078	2.01
4R	0.087	2.17	0.08	2.52
5R	0.079	3.34	0.079	4.3
6R	0.081	3.35	0.078	4.6
7R	0.074	1.71	0.077	2.45
Whole genome	0.078	2.2	0.078	2.77

Co-Localization of QTL

We identified 34 cross-validated SNPs in 33 protein-coding sequences associated with more than one trait (Supplementary Table 6). The number of cross-validated MTA ranged from four (GYD) to 16 (TGW). On average, each of the seven ‘Lo7’ pseudomolecules carry 4.6 of these gene models, 35.3% of the predicted rye genes mapped to chromosome 7R. Three protein-coding sequences were associated with four traits: next to the already described associations for SECCE7Rv1G0479810 (Figure 10), SECCE5Rv1G0347130, encoding a putative transmembrane protein, was associated with HDT, STC, TAX, and WAX, and SECCE7Rv1G0477980, encoding a cysteine-rich receptor-kinase-like protein, with PHT, TIN, STC, and TAX. Likewise, SECCE1Rv1G0052830 and SECCE7Rv1G0484030, encoding a methyltransferase and a cytochrome P450 family protein, were associated with TGW, GPC, and STC, while the predicted chitinase-encoding gene SECCE7Rv1G0521190 was associated with TGW, HDT, and WAX.

DISCUSSION

To fully exploit the recently published reference genome sequence data for rye (Li G. et al., 2021; Rabanus-Wallace et al., 2021), a systematic approach for the discovery of gene function is required. Genome-wide association mapping was hailed as the key gene discovery paradigm to translate the expectations connected to high-quality genome sequences into practice, providing the insights needed to develop better diagnostic, prognostic, and preventive strategies for different traits (Liu and Yan, 2019; Loos, 2020). Indeed, GWAS has become a central approach to studying the natural variation and mapping quantitative traits of cereals (Alqudah et al., 2019). However, in contrast to wheat and barley, target genotypes in outbreeding rye are highly heterozygous. This results in decreasing genotypic correlation between the line *per se* and test cross performance with the increasing complexity of a trait (Sprague and Tatum, 1942; Miedaner et al., 2014). We have, thus, studied two large multiple-hybrid populations to dissect the genetic architecture of important agronomic and quality traits in rye.

TABLE 5 | Survey on cross-validated marker/trait-associations for agronomic traits, plant height, heading date and grain quality traits mapped in the 'Lo7' genome assembly.

Lo7	PHT	HDT	GYD	TIN	TGW	GPC	STC	TAX	WAX
0R	3	7	10	3	7	4	4	4	6
1R	13	16	1	1	17	2	6	19	7
2R	28	13	3	1	11	6	15	8	4
3R	11	9	5	1	14	1	10	18	7
4R	14	8	3	2	5	2	1	8	11
5R	16	14	4	1	7	6	8	15	3
6R	11	31	11	0	18	9	6	2	9
7R	20	9	1	6	53	15	37	12	3
Σ	116	107	38	15	132	45	87	86	50
cds	35	31	13	9	32	13	27	31	15

GYD, grain yield; PHT, plant height; HDT, heading date; TIN, tiller number; TGW, thousand-grain weight; GPC, grain protein content; STC, starch content; TAX, total arabinoxylan content; WAX, water-extractable arabinoxylan content. 0R: unknown pseudomolecule; cds: protein coding sequence.

Prospects and Limits of Field Phenotyping in Rye

The rich genetic diversity in random mating rye populations has been depicted using DNA marker technologies (Hagenblad et al., 2016; Schreiber et al., 2018; Sidhu et al., 2019; Hawliczek et al., 2020). In the present study, we observed significant genotype variation ($p < 0.05$) for major target traits in a rye improvement program. The observed significant correlations reveal that complex target traits of rye breeding do not evolve independently. However, as weak and moderate correlations between traits dominate, the substitution of an inexpensive measurement for a correlated expensive phenotype in variety development is no constructive option. The strong to moderate negative correlation ($r = -0.9$ and $r = -0.39$) between GPC and STC in both data sets is well-known from studies in wheat (Muqaddasi et al., 2020; Yang et al., 2020). In contrast to previous research in biparental populations (Miedaner et al., 2012), the observed correlation between GPC and STC could be explained by the co-localization of QTL in our populations. The intermediate to high trait heritabilities (H^2) for GYD, PHT, and TGW compare well with previously published heritability estimates in biparental (Miedaner et al., 2012, 2018; Hackauf et al., 2017a) and multiple-hybrid populations (Auinger et al., 2016) of rye. This applies to STC (Miedaner et al., 2012) and HDT (Hackauf et al., 2017a) as well. The high heritability estimates for GPC, TAX, and WAX refer to high genetic variance and low error variance that is sufficiently powerful to detect additive gene variants even with minor effects. Heritability estimates of TIN compare well with data from wheat (Bilgrami et al., 2020) and uncovered a phenotyping bottleneck in our study. As we have studied TIN in both years in a large biparental population as well (Hackauf et al., 2017a) and due to a low throughput in assessing this trait (Wu et al., 2019; Reynolds et al., 2020), we were able to phenotype TIN in three environments only. At this point, the present study emphasizes that automated, nondestructive methods of phenotyping tiller traits at a high spatial resolution

and high throughput for large-scale assessment of small grain cereal accessions are urgently needed (Wu et al., 2019). The relevance of this conclusion is supported by previous research that identified genetic gains achieved in ear density, i.e., the number of effective tillers, as the main factor responsible for the progress in grain yield of hybrid rye cultivars (Laidig et al., 2017).

Novel Insights in Rye's Adaptation to Drought Stress

The atypical extreme climatic conditions in 2018 demonstrated that environmental impact is becoming increasingly relevant to agricultural production in Europe (Beillouin et al., 2020). Europe accounts for around 20% of the global cereal production (Schils et al., 2018) and is the main rye-producing area globally (FAOSTAT, 2020). Among the small grain winter cereals produced in Europe, rye revealed the lowest yield reduction compared to wheat, barley, and triticale when rainfall was fully excluded by means of rain-out shelters from tillering until harvest (Schittenhelm et al., 2014). The main driver enabling the performance of rye on light soils with low fertility and low water capacity is its highly developed root system. The entire root system of a single rye plant consisted of 13,815,672 branches, with a total length of 622 km, a surface area of 401 m², and a total root hair length of 11,000 km (Dittmer, 1937), which facilitates a very efficient uptake of water and nutrients (Paponov et al., 1999; Yeo et al., 2014). Irrespective of the powerful root system, average drought-induced grain yield reduction of 23.8% has been reported for hybrid rye in non-irrigated compared with the irrigated regime under natural drought stress conditions (Hübner et al., 2013), while up to 57% grain yield reduction was observed in controlled environments under different drought regimes (Kottmann et al., 2016). Notably, knowledge of physiological mechanisms of drought tolerance in rye is scarce (Hübner et al., 2013). As demonstrated on a large scale in the present study, natural drought stress conditions in Europe were detrimental in 2011 to grain yield of rye as compared with less-osmotic stress conditions in 2012. Our comprehensive phenotyping identified morphological adjustment of plant height and tiller number as drought stress responses of rye. Supported by a significant ($p < 0.001$) negative correlation between PHT and STC, the higher sink activity of grains under drought suggests a mobilization and reallocation of stem reserves to the grain, as has been observed in Sorghum (Blum et al., 1997). Interestingly, neither STC nor GPC was correlated with TGW under drought.

The significant opposite correlations between GYD and STC as well as PHT and STC in 2011 demonstrate that the large stem reserve storage in modern hybrid cultivars represents an adaptation strategy of rye for stable grain filling under osmotic stress. This protective measure needs to be considered when using the gibberellin (GA)-sensitive dwarfing gene *Ddw1* (Börner and Melz, 1988) as a breeder's option to improve plant height and lodging resistance in hybrid rye (Braun et al., 2019) and asks to search for an optimum in plant height of semi-dwarf rye hybrids. However, the significant negative correlations between PHT and STC in both years distinguish the potential of further improvements in PHT as a viable option to tap currently

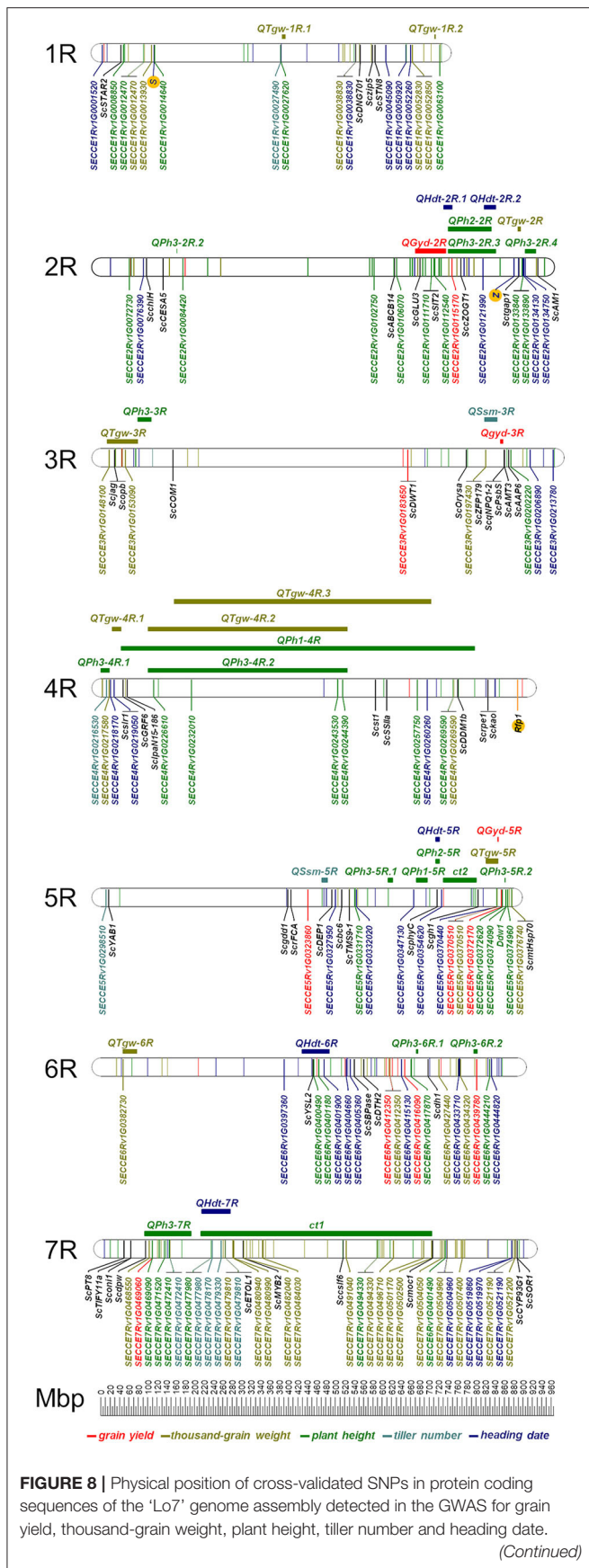
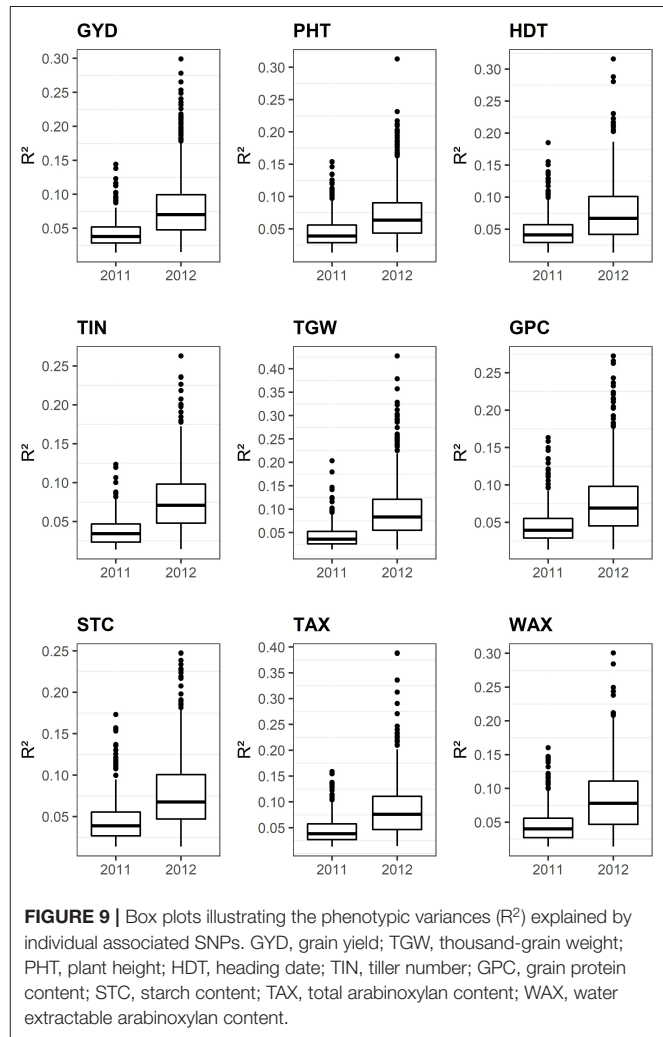


FIGURE 8 | The positions of both self-incompatibility loci, S and Z, the restorer-of-fertility locus *Rfp1* depicting the rye's unique reproduction biology, and the GA-sensitive dwarfing gene *Ddw1* are given as well. For 'Lo7' orthologs of cloned rice QTL the corresponding rice gene symbols were adapted to rye. The positions of the markers in the 'Lo7' physical map are given in Mbp. The horizontal bars and QTL symbols indicate the position of grain yield (QGyd-2R), heading date (QHdt), thousand-grain weight (QTgw), plant height (QPh), and spikes per squaremeter (QSsm).



unused potential for optimized dry matter allocation to the rye grain. Indeed, as no correlation between GYD and PHT was found in both years, this approach is not compromised by a relationship between grain yield and tall plant stature, as has been reported in a biparental population in rye (Miedaner et al., 2012). Significant ($p < 0.01$) positive correlations with PHT in both years indicate that HDT might be involved as a confounding variable. Although early heading date and a shorter vegetative phase were significantly ($p < 0.001$) correlated with grain yield under less-osmotic stress conditions in 2012, drought escape through earlier heading could not be identified as an adaptive

mechanism of rye in 2011, as indicated by a correlation coefficient between GYD and HDT of almost zero. Further research directed to the collection of drought tolerance traits that are not easy to measure under field conditions, and the assessment of rye using state-of-the-art phenotyping in controlled environments (Langstroff et al., 2021) is necessary to consolidate the relevance of our conclusion. The substantially increased grain arabinoxylan content in osmotically stressed rye is a striking phenotype that will subsequently be discussed.

High Levels of Genetic Diversity in Elite Rye Germplasm

The accessibility of distinct heterotic gene pools is the central pillar for the breeding of hybrid cultivars (Labroo et al., 2021). We investigated the extent of genetic differentiation, population clustering, and patterns of relationship among a diverse set of elite rye inbred lines based on SNP markers. All multivariate methods revealed the presence of two major groups, which was in perfect agreement with pedigree information, as all lines with similar pedigree clustered into the same group. The observed differences between inbred lines from the 'Petkus' and 'Carsten' pool concerning the average heterozygosity reflected that line development in the 'Carsten' pool is currently not going for complete homozygosity since a genetically broader synthetic composed of two less inbred lines is a more secure pollinator due to a longer pollen-shedding period (Geiger and Miedaner, 2009). We observed a high pairwise differentiation (F_{ST}) among the 'Petkus' and 'Carsten' populations, which coincides with the identified heterotic pattern in rye (Hepting, 1978) and indicates that these groups obviously followed different domestication and/or artificial selection paths. Our results generally agree with previous studies (Bauer et al., 2017; Vendelbo et al., 2020) that reported high genetic differentiation between parental populations of two other hybrid rye breeding programs as well. It is worth noting that the F_{ST} between elite 'Petkus' and 'Carsten' rye inbred lines estimated in the present study is higher than the differentiation ($F_{ST} = 0.10$) estimated from single nucleotide polymorphisms between elite and exotic wheat lines (Boeven et al., 2020). As the genetic diversity described in the analyzed rye germplasm has successfully been used to exploit commercial heterosis (Bundessortenamt, 2013), data presented in the present study may illustrate the (i.) sought genetic diversity and (ii.) command variable for marker-assisted management of elite germplasm pools in wheat hybrid breeding programs. Although comprehensive analyses have recently reported the genetic distance between parental populations as a crucial parameter to maximize the exploitation of heterosis for rye (Vendelbo et al., 2020) and wheat improvement (Boeven et al., 2020), it needs to be considered that heterotic response between genetically divergent groups cannot be predicted from genetic distances based on DNA markers but must be evaluated in field trials (Melchinger, 1999). Irrespective of that, SNP markers with high polymorphism information content (PIC) and Nei's gene diversity across 'Lo7' pseudomolecules may be useful to develop a subset of SNPs for routine studies where only a small to moderate number of SNPs are needed, as

is the case in mapping, marker-assisted recurrent selection, marker-assisted backcrossing, and quality control applications in rye.

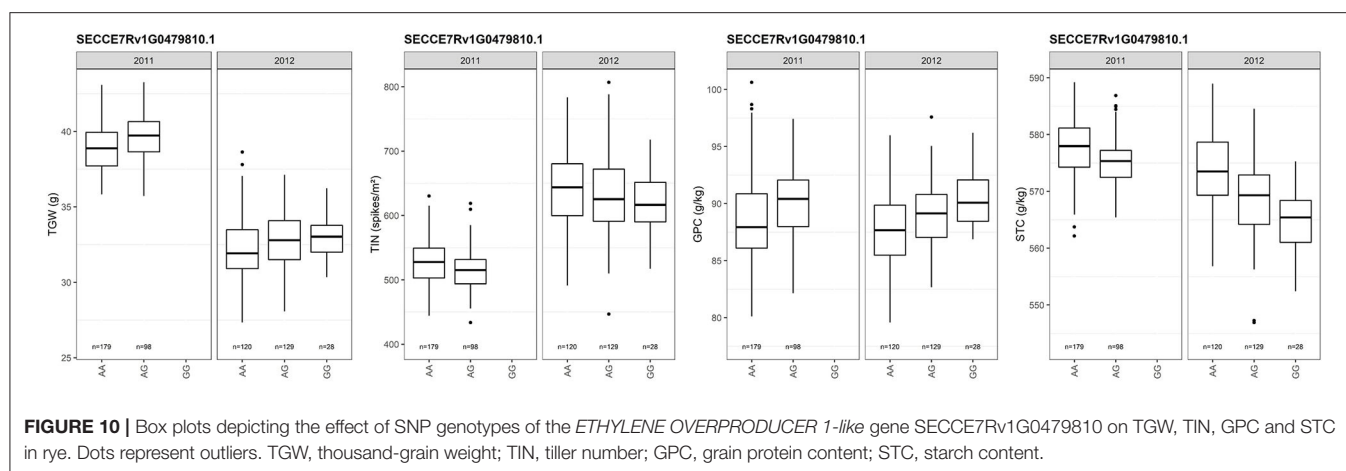
Rapid Decay of LD in Elite Rye Germplasm

Almost 99% of the SNP tags described in the present study are not represented by the Rye 600k SNP array (Bauer et al., 2017). The novel high-density map further extends previous versions established for the L2039-N x DH mapping population based on gene-based SNPs (Martis et al., 2013), as well as first-generation DArTTM markers (Hackauf et al., 2017a). The markers are evenly distributed through the 'Lo7' genome sequence, and the length of the linkage map is in the same order as other genetic maps in rye (Martis et al., 2013; Milczarski et al., 2016; Borzecka et al., 2018).

We observed low levels of LD extend across the entire genome in elite rye germplasm and in a similar order as most recently estimated for wheat (Liu et al., 2020). A rapid decline of LD fits to the expectation for an outcrossing species with low ancestral LD (Aunger et al., 2016). More recently, a higher mean genome-wide LD with several distinct blocks of strong LD has been observed in another rye germplasm (Vendelbo et al., 2020). This unexpected attribute refers to the use of the Gülzow (G) male-sterility system (Melz et al., 2003) as a hybridization system that obviously constitutes a population-determining parameter. Indeed, the G cytoplasm belongs to a plasmotype that is common in the Central European rye gene pool (Stojałowski et al., 2008). The spread in random-mating populations of cereal rye indicates that this plasmotype obviously provides a fitness advantage in female function [see Rieseberg and Blackman (2010) for review] that, in turn, resulted in a high frequency of restorer alleles (Łapiński and Stojałowski, 2003). This has a strong impact on practical applications. The unusual high LD reported by Vendelbo et al. (2020) demonstrates that, due to a low frequency of reliable maintainer genotypes, the development of male sterile seed parent lines for breeding of G-type CMS hybrid rye is a challenging task. In contrast, the CMS-inducing P cytoplasm used in the present study is easy to maintain and enables an unbiased capture and management of genetic diversity in the seed parent pool. The rapid rate of LD decay in P-type CMS hybrids promises a high resolution in GWAS and suggests to advance research on rye as a complement to barley, the model for the genetics and genomics of the Triticeae tribe (Han et al., 2020; Jayakodi et al., 2020) that features a comparatively higher mean LD (Malysheva-Otto et al., 2004; Comadran et al., 2009; Rode et al., 2012; Bellucci et al., 2017). As LD has been reported to decay rapidly within ~520 bp on average in rye (Li et al., 2011), follow-up experiments based on a higher SNP density than in the present study will enable to accurately predict LD beyond genetic maps.

Bridging the Genotype-Phenotype Gap in Rye

We integrated the novel SNP markers in the L2039-N x DH genetic map and the recently established rye reference genome. We obtained both genetic and physical positions for markers representing genetic diversity that is managed in the elite germplasm of a hybrid rye breeding program. The DArTTM



genetic map enabled the confident chromosomal localization of 90 SNP markers of unknown physical position in the ‘Lo7’ genome assembly to genome-specific regions and provides map position of 23 protein-coding genes. Both, the ‘Lo7’ reference genome sequence as well as the high-density map as a complement, will support marker development, marker-assisted selection (MAS), gene discovery and isolation, and, as demonstrated in the present study, GWAS. Furthermore, functional markers can explain a large part of the genetic variance, which may improve the predictive ability of genomic selection models in plant breeding programs (Liu et al., 2019). We ascertained GWA QTL that overlapped with previously identified QTL for plant height and agronomic traits, including grain yield (Hackauf et al., 2017a; Miedaner et al., 2018), by integrating the latter into the ‘Lo7’ assembly. Except for QTL residing on chromosome 4R, the physical distance covered by the QTL is low. It is currently unclear whether the QTL on chromosome 4R resides in a rarely recombining region, as has been demonstrated for the GA-insensitive dwarfing gene *compactum1* (*ct1*) mapping to a region with a high ratio of physical to a genetic distance of 122.4 Mbp/cM on chromosome 7R (Hackauf et al., 2021). Alternatively, the observation may owe to the many inversions and large structural rearrangements that have been observed among non-‘Lo7’ *Secale* genotypes relative to ‘Lo7’ (Rabanus-Wallace et al., 2021). Likewise, the large 4R intervals may be attributed to copy number variation for the QTL markers in the biparental mapping population that is not mirrored in the ‘Lo7’ genome assembly.

Anchoring the SNPs to the ‘Lo7’ reference genome is in the same order of magnitude as recently reported for the alignment of SNPs and SilicoDART™ markers to the hexaploid wheat RefSeq v1.0 reference genome (Sansaloni et al., 2020). The ‘Weining’ genome assembly (Li G. et al., 2021) offers a further resource to anchor additional rye DARTseq™ SNP tags with locations in physical space. Alignment of (i) DARTseq™ SNP tags representing protein-coding genes, as well as (ii) the corresponding protein-coding genes in rye and wheat, will support the identification of diversity that is associated with important agronomic and quality traits but that may be missing

in current wheat-improvement programs. As expected, 60–70% of the SNP markers were in intergenic regions, reflecting that genic regions are evolutionarily more conserved compared with intergenic regions, which evolve faster and accumulate higher levels of polymorphism. However, part of these SNPs may reside in promoters and regulatory elements and may represent functional markers as well.

The random-mating populations ‘Halo’ and ‘Carokurz’ as well as the two inbred lines ‘Lo7’ and ‘Lo225’ represent the heterotic gene pools ‘Petkus’ and ‘Carsten,’ respectively, and revealed significant differences in monoploid genome size (Rabanus-Wallace et al., 2021). This striking observation suggests the occurrence of structural variants (SVs) in the rye genome. Indeed, Hi-C asymmetry plots revealed SVs between the genomes of ‘Lo7’ and representatives of the genus *Secale*, including the inbred line ‘Lo225’ (Rabanus-Wallace et al., 2021). In the present study, DARTseq™ detected more than 27k SVs in the analyzed rye germplasm markers that could be mapped to the ‘Lo7’ genome assembly (data not shown). The impact of these SilicoDART™ on rye phenotypes deserves further research.

Controlling the Rate of False Discoveries

Control of the false discovery rate (FDR) in genetic association studies has become an increasingly appealing and accepted target for multiple-comparison adjustment (Brzyski et al., 2017). In order to control the false positive rate in multiple testing procedures, we have adjusted the level of statistical significance (p-value) of a single test so that the overall false control is still at a low level. While correction methods based on the family-wise error rate provide the most stringent control of false positives, we applied the FDR standard method according to Benjamini and Hochberg (1995) to better balance the false and hit and increase the power of GWAS (Zheng and Zhuo, 2017), as recently adopted in wheat (Ladejobi et al., 2019). In addition, we counterbalanced potential false positive associations due to statistical inferences (Liu and Yan, 2019) and other unaccounted factors, such as low-accuracy genotype calling at some loci (Browning and Yu, 2009) or small population size (Finno et al.,

2014) by cross-population and cross-species validation. The successful validation of candidate genes in the two different sets of experimental hybrids renders true associations more likely. This approach enabled to achieve high power in our rye GWAS to detect associations between SNPs and traits. The identified marker-trait associations reflect an appropriate sample size and substantial genetic diversity that determined the phenotypic variance of target traits in the studied elite germplasm. The rapid LD decay and rich genetic diversity of outbreeding species like rye are known to increase power in GWAS as compared with self-fertilizing species (Huang and Han, 2014). The GWAS reported in the present study serves as a foundation experiment by providing insights into the genetic architecture of important traits for rye improvement, allowing the targeted choice of parents for subsequent experiments, like candidate gene knock-out, over-expression, or genetic complementation, that are, in any case, indispensable to validate genes underlying the analyzed traits. Hybrid rye breeding offers a further and pragmatic cross-validation strategy of a single genomic locus with possible influence on the phenotype. The high-quality genome sequences of rye (Li G. et al., 2021; Rabanus-Wallace et al., 2021) enable the extension of sequence information for an efficient transformation of DArTseq™-based SNPs to single-plex Kompetitive Allele-Specific PCR (KASP) (Semagn et al., 2014) assays. These KASP markers will be used in progenies segregating for the superior (S) or inferior (I) SNP alleles to select and establish near-isogenic genotype bulks (NIB) of homozygous F3 lines, which will serve as pollinators in crosses with male-sterile single-cross testers between isolation walls (**Supplementary Figure 1**). The genetic makeup of these hybrids enables to precisely estimate phenotypic effects, recorded in target environments of rye production, as the difference ($\Delta S-I$) between the means of individual NIB partners, which either carry the S or the I allele at the candidate gene. This approach can, henceforth, take on central importance in ongoing efforts to isolate and characterize specific loci and bridge the genotype-phenotype gap for precision breeding in rye. However, it needs to be considered that the correlation between variants at a locus due to LD in experimental hybrids that have been established based on elite inbred lines further challenges the identification of causal variants, just as it is hard to find true signal overlaps between a GWAS and a QTL signal. Approaching target genes in gametes captured from random-mating rye populations with rapid decay of LD (Li et al., 2011) offers to overcome this challenge as well.

To conclude, the SNP catalog published with this paper can assist scientists to discover and use functional diversity in rye and related Triticeae species that may be essential for meeting the compulsions to act in modern agriculture under the directive of the multiple challenges concerning global food security (Lal, 2016; Liu et al., 2020).

An Extended View on the Genetic Basis of Variation for Complex Traits in Rye

Knowledge of genomic regions controlling complex traits is the key to our understanding of mechanisms behind trait architecture and for using them in marker-assisted crop

improvement programs. In the present study, both progeny sets of experimental rye hybrids from controlled crosses resulted in adequate statistical power to detect QTL, including those with small effects, and precisely map them in the 'Lo7' genome assembly. For all traits studied, a substantial proportion of the phenotypic variation can be explained with few loci of large effect, with the remainder due to numerous loci with small effect. QTL with large effects accounting for substantial proportions of phenotypic variance is well-known (Remington and Purugganan, 2003; Mackay, 2004; Roff, 2007) and has been identified in biparental rye mapping populations as well (Miedaner et al., 2012; Hackauf et al., 2017a). The large-effect QTL observed in the present study fit the model developed by Orr (1998), which suggests that natural selection validates mutations with large effects at the beginning of an adaptation process with a maximum of adaptive space, while, later on, in the process when the organism has essentially reached its optimum state, the space is narrowed and successful mutations must have smaller effects. In rye, an impressive selection gain for TGW from initially 19 g up to 57 g in a random mating population within two decades of breeding (Dill, 1983, 1989) supports the assumption that major genes controlling grain phenotypes exist. This conclusion is in line with high-heritability estimates and a preponderance of additive inheritance reported early on for TGW in rye (Wolski et al., 1972). Subsequently, major genes responsible for a substantial part of the heritable variance of grain weight in rye have been identified (Wricke, 2002; Skoryk et al., 2010). While the two complementary acting genes *Kernel weight 5* (KW5) and KW7 have been genetically (Wricke, 2002) and physically (Hackauf et al., 2021) mapped on chromosomes 5R and 7R, respectively, map positions of the genes *lg* (large grain) and *tg* (thick grain) identified by Skoryk et al. (2010) are unknown. It is worth noting in this context that there are seven major QTLs identified in the present study on chromosomes 2R, 4R, and 6R, each explaining more than 30% of the phenotypic variance for TGW. Interestingly, one of these QTL explaining 42.7% of the phenotypic variance is located just 3,443 bp upstream of the amino acid transporter SECCE4Rv1G0251940 on chromosome 4R. Most recently, overexpression of the amino acid transporter *TaAAP13* in wheat has significantly increased grain size, grain nitrogen concentration, and thousand-grain weight, indicating that the sink strength for nitrogen transport was increased by manipulation of amino acid transporters (Wan et al., 2021). The large-effect QTL, like those described for TGW and identified for GYD, HDT, GPC, STC, TAX, and WAX, are promising targets for successful marker-assisted selection in rye-improvement programs. Likewise, DNA markers derived from functionally characterized sequence motifs explaining part of the genetic variance have been shown to improve the predictive ability of genomic selection models in plant breeding (Spindel et al., 2016; Bian and Holland, 2017; Liu et al., 2019; Rice and Lipka, 2019). The present study reports SNP markers in candidate regions and genes for nine agronomic and quality traits of rye identified by GWAS methods. The distribution of the cross-validated SNPs indicates no accumulation of SNPs within short Mbp ranges, except for few loci governing TGW on 7R or

PHT on 2R. This observation is in line with the expectation for a cross-fertilizing species with a fast decline of LD.

We dissected trait correlations at the gene level and identified a subset of cross-validated SNP on protein-coding sequences associated with more than one trait. Genetic trait correlations might result either from pleiotropy or linkage disequilibrium. The predominant genetic basis of trait correlations is controversial and comprehensively reviewed by Chen and Lübberstedt (2010). The genetic diversity in the protein-coding sequences of the ‘Lo7’ genome assembly described in the present study provides candidate genes to further dissect the associated trait correlations based on dedicated genetic and genomic approaches.

Conserved Genetic Architecture for Complex Traits in Rye and Rice

Rice was the first sequenced crop genome, paving the way for the sequencing of additional and more complex genomes within the grass family (Jackson, 2016). Along with large-scale high-throughput genome-sequencing projects (Wang et al., 2018), rice genomics advanced our understanding of molecular mechanisms-controlling agronomic traits (Li et al., 2018; Song et al., 2018; Yao et al., 2018) and pioneered the direct transfer of basic research to field applications (Wang and Li, 2019; Wang et al., 2021). The common evolutionary origin of the grasses (Pont et al., 2019) served to make use of the rice genome sequence as a blueprint for marker development in rye (Hackauf and Wehling, 2005; Hackauf et al., 2012, 2017b). High-throughput transcript mapping, chromosome survey sequencing, and integration of conserved synteny information of model grass genomes identified 17 conserved syntenic linkage blocks, making up the rye genome in comparison to model grass genomes, including rice (Martis et al., 2013). In wheat and barley, yield-related genes have been identified based on their orthologous genes in rice [cv. Nadolska-Orczyk et al. (2017) for review]. Most recently, 237 orthologs of cloned rice QTL have been reported as candidate genes for yield and yield-related traits in a Meta-QTL (MQTL) analysis in bread wheat (Yang et al., 2021). BLASTP sequence similarity searches revealed that neither of the 237 wheat genes correspond to any of the ‘Lo7’ gene models associated with yield and yield-related traits in rye. This observation may be attributed to the relatively long linkage disequilibrium decay distance of wheat and a considered co-localization of associated markers obtained from GWAS and an MQTL within a 5-Mb physical region (Yang et al., 2021). The relevance of rice as a model crop for agronomic important traits in grasses is further emphasized by comparative genomic analyses that identified ortho-MQTL at co-linear regions between rice, barley, and maize, respectively (Khahani et al., 2020). The comparative analysis between rye and rice for similar or homologous traits conducted in the present study identified a conserved genetic architecture for agronomic traits that served as cross-species validation of individual MTAs. As a consequence, the commonality between the quantitative trait physiology and the biochemical function of a gene improves our understanding of the molecular nature of QTL in rye and extends our

knowledge about causal quantitative trait gene(s) (QTGs) in complex cereal genomes. Given a close evolutionary relationship among grass genomes (Pont et al., 2019), the genomic resources that have been developed for rye (Bauer et al., 2017; Li G. et al., 2021; Rabanus-Wallace et al., 2021) in combination with options of sophisticated experimental designs offered by hybrid rye breeding (Supplementary Figure 1) enable a systematic evaluation of the rich genetic diversity of rye in orthologs of cloned rice QTL for the discovery of gene function to further advance genomics-assisted Triticeae improvement.

SMART Breeding for Ergot Defense in Rye

The GWAS described in the present study offers novel options for the selection with markers and advanced reproductive technology (SMART) breeding (Davis et al., 1997) to promote the genetic improvement of rye in terms of high yield potential and minimized risk of ergot infestation. Because of the toxicity of ergot sclerotia for humans and animals, the European Commission (EC) Regulation (EU) 2021/1399 amending Regulation (EC) No. 1881/2006 further lowers maximum levels of ergot sclerotia and ergot alkaloids in rye and rye-milling products to 0.2 g/kg as from 1.7.2024. The inclusion of ergot reaction in the German national listing trials is attributed to the genetic diversity of winter rye cultivars in their susceptibility to ergot (Miedaner et al., 2010) and motivated the development of cultivars with improved ergot defense. CMS-based hybrids with an unsatisfactory restoration level and reduced pollen shedding are notably susceptible to ergot as the fungal spores have no competitors during the infection of the stigmatic tissue (Hackauf et al., 2012, 2017b). Restorer-of-fertility (*Rf*) genes are of central importance for cereal hybrid breeding, both for minimizing ergot infestation (Miedaner et al., 2010), as well as achieving maximum seed setting (Whitford et al., 2013). Indeed, P-type rye hybrids carrying an effective *Rfp* gene suffer from a significant reduction in grain yield (Miedaner et al., 2017). As a consequence of the high yield penalty, a restricted integration of restorer genes like *Rfp1* from weedy rye (Hackauf et al., 2012) in the pollinator gene pool, gaining a restorer index of ~50%, is considered as a feasible practice (Miedaner et al., 2017). However, it needs to be considered that rainy weather at a flowering time reduces pollen shedding and pollen movement. As wet pollen agglutinates and distributes over short distances only, a restorer index of ~50% may result in insufficient quantities of pollen to combat the fungus adequately. In order to comprehensively reduce the risk of ergot infection in hybrid rye, varieties should be developed with a restorer index of 100%, i.e., restoration of male fertility in every single plant of hybrid rye. This strategy is the key to support short-distance pollen distribution, as hybrid rye is able to set seeds upon self-pollination, just like wheat or barley. The associations between genetic causes of phenotypic variation in yield and yield components identified in ‘Lo7’ gene models in the present study enhance marker-assisted approaches to improve the ergot defense of rye that is currently solely focused on quick and accurate tracking of *Rfp* genes (Hackauf et al., 2012, 2017b). Knowledge of major QTL-controlling TGW like that residing in close proximity of the

amino acid-transporter SECCE4Rv1G0251940 offers a chance to precisely assess natural genetic diversity and counterbalance linkage drag effects of effective *Rfp* genes, as TGW counts among the traits negatively affected by *Rfp* genes (Miedaner et al., 2017). SNPs identified in ‘Lo7’ gene models encoding proteins with a crucial role in plant development like SECCE3Rv1G0201750 or SECCE7Rv1G0479810 provide valuable means for this purpose as well. A particularly attractive target to counterbalance fitness costs of *Rfp* genes measurable as inferior performance in GYD is SECCE3Rv1G0183650. SECCE3Rv1G0183650 is the ortholog of *OsDWT1*, a WUSCHEL-related homeobox (WOX) transcription factor that promotes tiller growth downstream of *SLR1* in rice (Wang et al., 2014). The number of tillers produced per plant is controlled by the environment during the period of tiller development from the three-leaf stage to jointing and the amount of tiller mortality that occurs from jointing to anthesis (Shaaf et al., 2019; Tilley et al., 2019). Recently, empirical data have expanded our understanding of the physiological mechanisms underpinning the yield response to plant density. While a high tillering potential reduces the agronomic optimum plant density in both high and low yield environments, at per-plant scale, a compensation between heads per plant and kernels per head was the main factor contributing to yield with different tillering potentials under varying yield environments (Bastos et al., 2020). Knowledge of genes like SECCE3Rv1G0183650 and the cross-validation strategy described before (Supplementary Figure 1) further advances our knowledge of this critical yield component in order to develop rye varieties with an optimal and environmentally stable tillering potential.

Advancing Rye to an Authentic High-Performance Crop With Diverse End-Use

Arabinoxylans are non-starch polysaccharides and the predominant components within the endosperm cell walls in rye and, to a lesser degree, in wheat (Bukša et al., 2016; Freeman et al., 2016; Oest et al., 2020). High AX content increases the falling number, dough yield, bread volume, and bread shelf life (Bukša et al., 2010; Oest et al., 2020). Current methods of rye breeding and the growth under severe drought conditions in a changing climate are thought to negatively influence bread qualities, which demands improved understanding of the mechanisms by which proteins, starch, and AX—the most prominent hemicelluloses—might interact (Oest et al., 2020). To increase the value of rye as livestock feed, a low-WAX content is currently considered as a desired grain phenotype (Kobylyansky et al., 2019), in sharp contrast to the optimal needs for bread making (Bukša et al., 2010; Oest et al., 2020). Interestingly, recent research has provided a novel momentum concerning the value of rye AXs for pig feeding. Indeed, beneficial changes in the physicochemical characteristics of digesta of young pigs due to increased rye levels in the diet have been attributed to the very high content of AXs as the predominant “dietary fiber” content of rye, which is beneficial for improving “gut health,” an important parameter in terms of animal health, animal welfare, and food safety (Wilke et al., 2021). In any case, the evaluation of

end-use quality parameters like WAX asks for molecular markers that are currently not available for large-scale genotyping of WAX in rye. Our cross-species validation did neither for TAX nor WAX result in the identification of obvious candidate genes. This observation might refer to higher selection pressure on genes controlling these grain-quality parameters in rye or a lack of synteny in rye and rice. However, cross-population validation identified 31 and 15 protein-coding genes, respectively, that deserve a more detailed examination. Remarkably, more than 30% of the cross-validated protein-coding genes associated with WAX are predicted to encode protein kinases, including receptor protein kinases. Receptor protein kinases are discussed to sense cell-wall perturbations originating from osmotic stress (Zhu, 2016). Stress upregulates the expression of expansins and xyloglucan-modifying enzymes that can remodel cell walls (Tenhaken, 2015). Notably, we identified an association of SECCE6Rv1G0407620, encoding a Xyloglucan alpha-1,6-xylosyltransferase and WAX. Most interestingly, we identified major effects ($R^2 > 10\%$) of a SNP in the receptor protein kinase-like gene model SECCE4Rv1G0223400 on WAX. This observation supports previous research highlighting the impact of receptor protein kinases function in stress responses (Marshall et al., 2012). Further studies will demonstrate if SECCE4Rv1G0223400 is the supposed key regulator of WAX that contributes to improving performance of rye under drought stress. Likewise, in-depth analysis of SECCE4Rv1G0223400 for functional SNPs controlling WAX content depicts an innovative example for SMART breeding of high-quality feed rye varieties with stable contents of WAX. The identification of natural genetic diversity controlling these grain quality traits is particularly important for commercial rye breeding, as the ability to substantially increase TAX and WAX appears a crucial adaptation strategy to drought stress in rye. The knowledge gained in the present study is consistent with previous research in wheat, reporting on increased concentration of AX upon drought stress (Hong et al., 1989; Coles et al., 1997; Gebruers et al., 2010; Rakszegi et al., 2014). AXs are the dominant non-cellulosic polysaccharides in the thick aleurone cell walls in cereal grains, and the second most abundant component in the starchy endosperm cell walls after (1,3;1,4)- β -glucan (Rosicka-Kaczmarek et al., 2016; Hassan et al., 2017). The increase in the dietary fiber AX in rye and wheat under drought stress conditions contributes to remodeling the cell wall composition as a strategy in response to abiotic stress (Tenhaken, 2015). In barley and wheat, AX constitutes 4.2–9.6% and 4.1–9% of grain dry matter, respectively (Martinant et al., 1999; Izydorczyk and Dexter, 2008; Andersson et al., 2009). The effective tolerance of rye toward drought stress (Schittenhelm et al., 2014) is mirrored by the largest amounts of grain AX among cultivated Triticeae species, ranging from 8 to 12.1% (Rosicka-Kaczmarek et al., 2016). As efficient responses to purifying selection as well as significant genetic gains in agronomic traits were feasible not before hybrid breeding started 50 years ago (Laidig et al., 2017), high AX content of rye grains could evolve due to long-lasting natural selection of random mating rye populations in harsh European environments north of the alps on poor, podsolic soils. The significant ($p < 0.001$) negative correlation between

GYD and WAX in 2011 refers to a changed carbon allocation and metabolism that resulted in energy dissipation in terms of declined yield. Recent progress in developing random mating feeding rye with low content of WAX (Kobylyansky et al., 2019) has demonstrated that this relationship can be overcome by breeding. For the selection of genotypes with low content of WAX in Central European rye breeding programs, line *per se* performance in well-defined drought stress conditions of rain-out shelters appears a promising strategy due to lower costs of seed production, the higher selection intensity, and the larger proportion of additive genetic variance exploited in inbred lines as compared to hybrids (Miedaner et al., 2014). The significant but weak phenotypic correlation $r_p = 0.25\text{--}0.28$ between line *per se* and testcross performance reported for WAX (Miedaner et al., 2014) probably refers to the different environmental conditions of field experiments conducted by Miedaner and co-workers in 2010 and 2011 and should, thus, not compromise the proposed selection strategy. Levels of precipitation were higher, and average temperature was lower in 2010 as compared to the long-term mean [DWD (Deutscher Wetterdienst), 2010] and the already described natural drought stress in 2011. The cross-validated SNP markers identified in the present study provide essential targets for further research to overcome the strong impact of the environment on TAX and WAX by an efficient and accurate selection of suitable genotypes for the development of rye with certified grain qualities.

Conclusions

Despite formidable achievements, major challenges in rye production remain. Breeding of cultivars with high yield potential, strong ergot defense, and tailor-made grain qualities is inevitable to further advance rye from an all-rounder to an authentic high-performance crop with different and certified types of end-use. For this purpose, further progress in rye phenomics and functional genomics research is necessary to associate genome sequence information with phenotypes related to rye growth and development. The present study reports candidate regions and genes in the recently published ‘Lo7’ high-quality genome assembly (Rabanus-Wallace et al., 2021) for nine agronomic and quality traits of rye identified by GWAS methods as a crucial step to make previously hidden genetic variation accessible to genetic studies and breeding of rye. The observed rich genetic diversity of elite rye germplasm, together with a bulked segregant phenotyping strategy of testcross performance in multi-environmental field trials, supports previous arguments (Hackauf et al., 2017a) for a stronger utilization of rye in research directed to the identification of valuable alleles for Triticeae improvement programs. To conclude, the genomic data generated in this study improve our understanding of the allelic variation in rye germplasm collections and will facilitate the advancing of genomics-assisted rye breeding for variety improvement as well.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The work presented here was carried out in collaboration between all the authors. BH and FF defined the research theme. BH conceived the design of this study, coordinated the experiments, supervised the project, and performed comparative analyses. FF developed plant materials, conducted field trials, and performed the phenotyping of agronomic traits, including grain yield. GJ assessed grain-quality parameters. AK conducted genotyping. BH and DS established the genetic linkage map. DS conducted the GWAS and the analyses of population genetic data. BH, DS, AK, and AZ discussed analyses and interpretation of results. BH wrote and DS, AZ, as well as AK, reviewed and edited the manuscript. All the authors have read and approved the final version of the manuscript.

FUNDING

This study was financially supported by the German Federal Ministry of Food and Agriculture based on the decision of the Parliament of the Federal Republic of Germany through the Fachagentur Nachwachsende Rohstoffe e. V. (Grant Nos. 22014509 and 22005910). We gratefully acknowledge support by the Open Access Publication Funds of the Julius Kühn-Institute.

ACKNOWLEDGMENTS

We would like to thank both reviewers and the handling editor at Frontiers in Plant Science for their critical reading and valuable feedback on the manuscript. We gratefully acknowledge the excellent technical assistance of Gunda Kölzow and Rita Friedrich. Sincere thanks are given to the teams at the respective locations for their excellent technical support and to Magdalena Góralska for helpful comments on the manuscript. This paper is dedicated to Prof. Dr. agr. Dr. h.c. Hartwig H. Geiger. His fundamental research on hybrid rye breeding enabled the present study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.718081/full#supplementary-material>

Supplementary Figure 1 | Flow diagram illustrating seed production of near-isogenic experimental rye hybrids for cross-validation of candidate genes by bulked segregant phenotyping. P1, P2: parental genotypes, M, m: superior and inferior allele at a given SNP marker.

REFERENCES

- Agelet, L. E., and Hurburgh, C. R. Jr. (2010). A tutorial on near infrared spectroscopy and its calibration. *Crit. Rev. Anal. Chem.* 40, 246–260. doi: 10.1080/10408347.2010.515468
- Alqudah, A. M., Sallam, A., Baenziger, S. P., and Börner, A. (2019). GWAS: fast-forwarding gene identification and characterization in temperate Cereals: lessons from Barley - a review. *J. Adv. Res.* 22, 119–135. doi: 10.1016/j.jare.2019.10.013
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Andersson, R., Fransson, G., Tietjen, M., and Åman, P. (2009). Content and molecular-weight distribution of dietary fiber components in whole-grain rye flour and bread. *J. Agric. Food Chem.* 57, 2004–2008. doi: 10.1021/jf801280f
- Auinger, H. J., Schönleben, M., Lehermeier, C., Schmidt, M., Korzun, V., Geiger, H. H., et al. (2016). Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye (*Secale cereale* L.). *Theor. Appl. Genet.* 129, 2043–2053. doi: 10.1007/s00122-016-2756-5
- Bastos, L. M., Carciocchi, W., Lollato, R. P., Jaenisch, B. R., Rezende, C. R., Schwalbert, R., et al. (2020). Winter wheat yield response to plant density as a function of yield environment and tillering potential: a review and field studies. *Front. Plant Sci.* 11:54. doi: 10.3389/fpls.2020.00054
- Bauer, E., Schmutzer, T., Barilar, I., Mascher, M., Gundlach, H., Martis, M. M., et al. (2017). Towards a whole-genome sequence for rye (*Secale cereale* L.). *Plant J.* 89, 853–869. doi: 10.1111/tpj.13436
- Beillouin, D., Schauburger, B., Bastos, A., Ciais, P., and Makowski, D. (2020). Impact of extreme weather conditions on European crop production in 2018. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 375:20190510. doi: 10.1098/rstb.2019.0510
- Bellucci, A., Tondelli, A., Fangel, J. U., Torp, A. M., Xu, X., Willats, W. G., et al. (2017). Genome-wide association mapping in winter barley for grain yield and culm cell wall polymer content using the high-throughput CoMPP technique. *PLoS ONE* 12:e0173313. doi: 10.1371/journal.pone.0173313
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Statistical Soc.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bevan, M. W., Uauy, C., Wulff, B. B., Zhou, J., Krasileva, K., and Clark, M. D. (2017). Genomic innovation for crop improvement. *Nature* 543, 346–354. doi: 10.1038/nature22011
- Bian, Y., and Holland, J. B. (2017). Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. *Heredity* 118:585. doi: 10.1038/hdy.2017.4
- Bilgrami, S. S., Ramandi, H. D., Shariati, V., Razavi, K., Tavakol, E., Fakheri, B. A., et al. (2020). Detection of genomic regions associated with tiller number in Iranian bread wheat under different water regimes using genome-wide association study. *Sci. Rep.* 10:14034. doi: 10.1038/s41598-020-69442-9
- Blum, A., Golan, G., Mayer, J., and Simmena, B. (1997). The effect of dwarfing genes on sorghum grain filling from remobilized stem reserves, under stress. *Field Crops Res.* 52, 43–54. doi: 10.1016/S0378-4290(96)03462-4
- Boeven, P., Zhao, Y., Thorwarth, P., Liu, F., Maurer, H. P., Gils, M., et al. (2020). Negative dominance and dominance-by-dominance epistatic effects reduce grain-yield heterosis in wide crosses in wheat. *Sci. Adv.* 6:eay4897. doi: 10.1126/sciadv.aay4897
- Bolibok-Bragoszewska, H., Heller-Uszyńska, K., Wenzl, P., Uszyński, G., Kilian, A., and Rakoczy-Trojanowska, M. (2009). DArT markers for the rye genome - genetic diversity and mapping. *BMC Genomics* 10:578. doi: 10.1186/1471-2164-10-578
- Börner, A., and Melz, G. (1988). Response of rye genotypes differing in plant height to exogenous gibberellic acid application. *Arch. Züchtungsforsch* 18, 71–74.
- Borzecka, E., Hawliczek-Strulak, A., Bolibok, L., Gawroński, P., Tofil, K., and Milczarski, P. (2018). Effective BAC clone anchoring with genotyping-by-sequencing and Diversity Arrays Technology in a large genome cereal rye. *Sci. Rep.* 8:8428. doi: 10.1038/s41598-018-26541-y
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Braun, E. M., Tsvetkova, N., Rotter, B., Siekmann, D., Schwefel, K., Krezdorn, N., et al. (2019). Gene expression profiling and fine mapping identifies a gibberellin 2-oxidase gene co-segregating with the dominant dwarfing gene *Ddw1* in Rye (*Secale cereale* L.). *Front. Plant Sci.* 10:857. doi: 10.3389/fpls.2019.00857
- Brinton, J., Ramirez-Gonzalez, R. H., Simmonds, J., Wingen, L., Orford, S., and Griffiths, S. (2020). A haplotype-led approach to increase the precision of wheat breeding. *Commun Biol.* 3:712. doi: 10.1038/s42003-020-01413-2
- Browning, B. L., and Yu, Z. (2009). Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *Am. J. Hum. Genet.* 85, 847–861. doi: 10.1016/j.ajhg.2009.11.004
- Brzyski, D., Peterson, C. B., Sobczyk, P., Candès, E. J., Bogdan, M., and Sabatti, C. (2017). Controlling the rate of GWAS false discoveries. *Genetics* 205, 61–75. doi: 10.1534/genetics.116.193987
- Buksa, K., Nowotna, A., Praznik, W., Gambus, H., Ziobro, R., and Krawontka, J. (2010). The role of pentosans and starch in baking of wholemeal rye bread. *Food Res. Intern.* 43, 2045–2051. doi: 10.1016/j.foodres.2010.06.005
- Buksa, K., Praznik, W., Loeppert, R., and Nowotna, A. (2016). Characterization of water and alkali extractable arabinoxylan from wheat and rye under standardized conditions. *J. Food Sci. Technol.* 53, 1389–1398. doi: 10.1007/s13197-015-2135-2
- Bundessortenamt (2013). *Beschreibende Sortenliste Getreide, Mais, Öl- und Faserpflanzen, Leguminosen, Rüben und Zwischenfrüchte 2013*. Hannover.
- Cavalli-Sforza, L. L., and Edwards, A. W. (1967). Phylogenetic analysis. Models and estimation procedures. *Am. J. Hum. Genet.* 19, 233–257.
- Chen, Y., and Lübberstedt, T. (2010). Molecular basis of trait correlations. *Trends Plant Sci.* 15, 454–461. doi: 10.1016/j.tplants.2010.05.004
- Coles, G. D., Hartunian-Sowa, S. M., Jamieson, P. D., Hay, A. J., Atwell, W. A., and Fulcher, R. G. (1997). Environmentally-induced variation in starch and non-starch polysaccharide content in wheat. *J. Cereal Sci.* 26, 47–54. doi: 10.1006/jcrs.1996.0102
- Comadran, J., Thomas, W. T., van Eeuwijk, F. A., Ceccarelli, S., Grando, S., Stanca, A. M., et al. (2009). Patterns of genetic diversity and linkage disequilibrium in a highly structured *Hordeum vulgare* association-mapping population for the Mediterranean basin. *Theor. Appl. Genet.* 119, 175–187. doi: 10.1007/s00122-009-1027-0
- Davis, G. P., D'occhio, M. J., and Hetzel, D. J. S. (1997). “SMART breeding: selection with markers and advanced reproductive technologies,” in *Proceedings of the 12th Conference of the Association for the Advancement of Animal Breeding and Genetics* (Dubbo, NSW), 429–432.
- Dill, P. (1983). Zur züchterischen Verbesserung der Kornmasse bei Winterroggen (*Secale cereale* L.). *Arch. Züchtungsforsch* 13, 157–168.
- Dill, P. (1989). Zur züchterischen Verbesserung der Kornmasse bei Winterroggen (*Secale cereale* L.) – Ergebnisse von Drillprüfungen. *Arch. Züchtungsforsch* 20, 329–337.
- Dittmer, H. J. (1937). A quantitative study of the roots and root hairs of a winter rye plant (*Secale cereale*). *Am. J. Bot.* 24, 417–420. doi: 10.1002/j.1537-2197.1937.tb09121.x
- DWD (Deutscher Wetterdienst). (2010). *Jahresrückblick: Deutschlandwetter im Jahr 2010*. Available online at: https://www.dwd.de/DE/presse/pressemitteilungen/DE/2010/20101229_deutschlandwetter_jahr.pdf?__blob=publicationFile&v=3
- DWD (Deutscher Wetterdienst) (2012). *Drought Conditions in Europe in the Spring of 2012*. Available online at: https://www.dwd.de/EN/ourservices/specialevents/drought/20120810_Trockenheit_2012_en.pdf?__blob=publicationFile&v=4
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491. doi: 10.1093/genetics/131.2.479
- FAOSTAT (2020). *Production. Crops. “Rye,” “World.”* Rome: FAO. Available online at: <http://www.fao.org/faostat/en/#data/QC> (accessed August 25, 2021).
- Finno, C. J., Aleman, M., Higgins, R. J., Madigan, J. E., and Bannasch, D. L. (2014). Risk of false positive genetic associations in complex traits

- with underlying population structure: a case study. *Vet. J.* 202, 543–549. doi: 10.1016/j.tvjl.2014.09.013
- Freeman, J., Lovegrove, A., Wilkinson, M. D., Saulnier, L., Shewry, P. R., and Mitchell, R. A. (2016). Effect of suppression of arabinoxylan synthetic genes in wheat endosperm on chain length of arabinoxylan and extract viscosity. *Plant Biotechnol. J.* 14, 109–116. doi: 10.1111/pbi.12361
- Gaikpa, D. S., Koch, S., Fromme, F. J., Siekmann, D., Würschum, T., and Miedaner, T. (2020). Genome-wide association mapping and genomic prediction of Fusarium head blight resistance, heading stage and plant height in winter rye (*Secale cereale*). *Plant Breed.* 139, 508–520. doi: 10.1111/pbr.12810
- Gao, L., Meng, C., Yi, T., Xu, K., Cao, H., Zhang, S., et al. (2021). Genome-wide association study reveals the genetic basis of yield- and quality-related traits in wheat. *BMC Plant Biol.* 21:144. doi: 10.1186/s12870-021-02925-7
- Gebruers, K., Dornez, E., Bedö, Z., Rakszegi, M., Frás, A., Boros, D., et al. (2010). Environment and genotype effects on the content of dietary fiber and its components in wheat in the HEALTHGRAIN diversity screen. *J. Agric. Food Chem.* 58, 9353–9361. doi: 10.1021/jf100447g
- Geiger, H. H., and Miedaner, T. (2009). “Rye breeding,” in *Cereals, Handbook of Plant Breeding*, Vol. 3, ed M. J. Carona (New York, NY: Springer Science + Business Media), 157–181.
- Geiger, H. H., and Schnell, F. W. (1970). Cytoplasmic male sterility in rye (*Secale cereale* L.). *Crop Sci.* 10:590–593. doi: 10.2135/cropsci1970.0011183X001000050043x
- Gruner, P., Schmitt, A. K., Flath, K., Piepho, H. P., and Miedaner, T. (2021). Mapping and validating stem rust resistance genes directly in self-incompatible genetic resources of winter rye. *Theor. Appl. Genet.* 134, 1989–2003. doi: 10.1007/s00122-021-03800-7
- Hackauf, B., Bauer, E., Korzun, V., and Miedaner, T. (2017b). Fine mapping of the restorer gene *Rfp3* from an Iranian primitive rye (*Secale cereale* L.). *Theor. Appl. Genet.* 130, 1179–1189. doi: 10.1007/s00122-017-2879-3
- Hackauf, B., Haffke, S., Fromme, F. J., Roux, S. R., Kusterer, B., Musmann, D., et al. (2017a). QTL mapping and comparative genome analysis of agronomic traits including grain yield in winter rye. *Theor. Appl. Genet.* 130, 1801–1817. doi: 10.1007/s00122-017-2926-0
- Hackauf, B., Korzun, V., Wortmann, H., Wilde, P., and Wehling, P. (2012). Development of conserved ortholog set markers linked to the restorer gene *Rfp1* in rye. *Mol. Breed.* 30, 1507–1518. doi: 10.1007/s11032-012-9736-5
- Hackauf, B., Rabanus-Wallace, M. T., and Korzun, V. (2021). “Bridging the genotype-phenotype gap for precision breeding in rye,” in *The Rye Genome, 1st Edn*, eds N. Stein, M. T. Rabanus-Wallace, and C. Kole (Compendium of Plant Genomes; Springer International Publishing), 46. Available online at: <https://www.springer.com/de/book/9783030833824>
- Hackauf, B., Rudd, S., van der Voort, J. R., Miedaner, T., and Wehling, P. (2009). Comparative mapping of DNA sequences in rye (*Secale cereale* L.) in relation to the rice genome. *Theor. Appl. Genet.* 118, 371–384. doi: 10.1007/s00122-008-0906-0
- Hackauf, B., and Wehling, P. (2005). Approaching the self-incompatibility locus Z in rye (*Secale cereale* L.) via comparative genetics. *Theor. Appl. Genet.* 110, 832–845. doi: 10.1007/s00122-004-1869-4
- Hagenblad, J., Oliveira, H. R., Forsberg, N. E., and Leino, M. W. (2016). Geographical distribution of genetic diversity in *Secale* landrace and wild accessions. *BMC Plant Biol.* 16:23. doi: 10.1186/s12870-016-0710-y
- Han, Y., Broughton, S., Liu, L., Zhang, X. Q., Zeng, J., He, X., et al. (2020). Highly efficient and genotype-independent barley gene editing based on anther culture. *Plant Commun.* 2:100082. doi: 10.1016/j.xplc.2020.100082
- Hardy, O. J. (2003). Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. *Mol. Ecol.* 12, 1577–1588. doi: 10.1046/j.1365-294X.2003.01835.x
- Hardy, O. J., and Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 2, 618–620. doi: 10.1046/j.1471-8286.2002.00305.x
- Hassan, A. S., Houston, K., Lahnstein, J., Shirley, N., Schwerdt, J. G., Gidley, M. J., et al. (2017). A Genome Wide Association Study of arabinoxylan content in 2-row spring barley grain. *PLoS ONE* 12:e0182537. doi: 10.1371/journal.pone.0182537
- Hawliczek, A., Bolibok, L., Tofil, K., Borzecka, E., Jankowicz-Cieślak, J., Gawroński, P., et al. (2020). Deep sampling and pooled amplicon sequencing reveals hidden genic variation in heterogeneous rye accessions. *BMC Genomics* 21:845. doi: 10.1186/s12864-020-07240-3
- Hepting, L. (1978). Analyse eines 7x7-Sortendiallels zur Ermittlung geeigneten Ausgangsmaterials für die Hybridzüchtung bei Roggen. *Z. Pflanzenzüchtung* 80, 188–197.
- Hill, C. B., Angessa, T. T., Zhang, X. Q., Chen, K., Zhou, G., Tan, C., et al. (2021). A global barley panel revealing genomic signatures of breeding in modern Australian cultivars. *Plant J.* 106, 419–434. doi: 10.1111/tpj.15173
- Hong, B. H., Rubenthaler, G. L., and Allen, R. E. (1989). Wheat pentosans. I. Cultivar variation and relationship to kernel hardness. *Cereal Chem.* 66, 369–373.
- Huang, X., and Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Ann. Rev. Plant Biol.* 65, 531–551. doi: 10.1146/annurev-arplant-050213-035715
- Hübner, M., Wilde, P., Schmiedchen, B., Dopierala, P., Gowda, M., Reif, J. C., et al. (2013). Hybrid rye performance under natural drought stress in Europe. *Theor. Appl. Genet.* 126, 475–482. doi: 10.1007/s00122-012-1994-4
- Izydorczyk, M. S., and Dexter, J. E. (2008). Barley β -glucans and arabinoxylans: molecular structure, physicochemical properties, and uses in food products - a review. *Food Res. Intern.* 41, 850–868. doi: 10.1016/j.foodres.2008.04.001
- Jackson, S. A. (2016). Rice: the first crop genome. *Rice* 9:14. doi: 10.1186/s12284-016-0087-4
- Jamann, T. M., Balint-Kurti, P. J., and Holland, J. B. (2015). QTL mapping using high-throughput sequencing. *Methods Mol. Biol.* 1284, 257–285. doi: 10.1007/978-1-4939-2444-8_13
- Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V. S., Gundlach, H., Monat, C., et al. (2020). The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* 588, 284–289. doi: 10.1038/s41586-020-2947-8
- Jürgens, H. U., Jansen, G., and Wegener, C. B. (2012). Characterisation of several rye cultivars with respect to arabinoxylans and extract viscosity. *J. Agricult. Sci.* 4, 1–12. doi: 10.5539/jas.v4n5p1
- Kempthorne, O. (1956). The theory of the diallel cross. *Genetics* 41, 451–459. doi: 10.1093/genetics/41.4.451
- Khahani, B., Tavakol, E., Shariati, V., and Fornara, F. (2020). Genome wide screening and comparative genome analysis for Meta-QTLs, ortho-MQTLs and candidate genes controlling yield and yield-related traits in rice. *BMC Genomics* 21:294. doi: 10.1186/s12864-020-6702-1
- Kobylyansky, V. D., Kuznetsova, L. I., Solodukhina, O. V., Lavrentyeva, N. S., and Timina, M. A. (2019). Prospects of using low-pentosan grain fodder rye for baking purposes. *Russ. Agricult. Sci.* 45, 1–4. doi: 10.3103/S1068367419010063
- Kottmann, L., Wilde, P., and Schittenhelm, S. (2016). How do timing, duration, and intensity of drought stress affect the agronomic performance of winter rye? *Europ. J. Agron.* 75, 25–32. doi: 10.1016/j.eja.2015.12.010
- Labroo, M. R., Studer, A. J., and Rutkoski, J. E. (2021). Heterosis and hybrid crop breeding: a multidisciplinary review. *Front. Genet.* 12:643761. doi: 10.3389/fgene.2021.643761
- Ladejobi, O., Mackay, I. J., Poland, J., Praud, S., Hibberd, J. M., and Bentley, A. R. (2019). Reference genome anchoring of high-density markers for association mapping and genomic prediction in European winter wheat. *Front. Plant Sci.* 10:1278. doi: 10.3389/fpls.2019.01278
- Laidig, F., Piepho, H. P., Rentel, D., Drobek, T., Meyer, U., and Huesken, A. (2017). Breeding progress, variation, and correlation of grain and quality traits in winter rye hybrid and population varieties and national on-farm progress in Germany over 26 years. *Theor. Appl. Genet.* 30, 981–998. doi: 10.1007/s00122-017-2865-9
- Lal, R. (2016). Global food security and nexus thinking. *J. Soil Water Conserv.* 71, 85A–90A. doi: 10.2489/jswc.71.4.85A
- Langstroff, A., Heuermann, M. C., Stahl, A., and Junker, A. (2021). Opportunities and limits of controlled-environment plant phenotyping for climate response traits. *Theor. Appl. Genet.* doi: 10.1007/s00122-021-03892-1. [Epub ahead of print].
- Lapiński, M., and Stojalowski, S. (2003). Occurrence and genetic identity of male sterility-inducing cytoplasm in rye (*Secale* spp.). *Plant Breed. Seed Sci.* 48, 7–23.
- Li, G., Wang, L., Yang, J., He, H., Jin, H., Li, X., et al. (2021). A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. *Nat. Genet.* 53, 574–584. doi: 10.1038/s41588-021-00808-z
- Li, M., Geng, L., Xie, S., Wu, D., Ye, L., and Zhang, G. (2021). Genome-wide association study on total starch, amylose and amylopectin in barley grain

- reveals novel putative alleles. *Int. J. Mol. Sci.* 22:553. doi: 10.3390/ijms22020553
- Li, Y., Haseneyer, G., Schön, C. C., Ankerst, D., Korzun, V., Wilde, P., et al. (2011). High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *BMC Plant Biol.* 11:6. doi: 10.1186/1471-2229-11-6
- Li, Y., Xiao, J., Chen, L., Huang, X., Cheng, Z., Han, B., et al. (2018). Rice functional genomics research: past decade and future. *Mol. Plant.* 11, 359–380. doi: 10.1016/j.molp.2018.01.007
- Liu, C., Sukumaran, S., Jarquin, D., Crossa, J., Dreisigacker, S., Sansaloni, C., et al. (2020). Comparison of array- and sequencing-based markers for genome-wide association mapping and genomic prediction in spring wheat. *Crop Sci.* 60, 211–225. doi: 10.1002/csc2.20098
- Liu, H. J., and Yan, J. (2019). Crop genome-wide association study: a harvest of biological relevance. *Plant J.* 97, 8–18. doi: 10.1111/tpj.14139
- Liu, X., Wang, H., Hu, X., Li, K., Liu, Z., Wu, Y., et al. (2019). Improving genomic selection with quantitative trait loci and nonadditive effects revealed by empirical evidence in maize. *Front. Plant Sci.* 10:1129. doi: 10.3389/fpls.2019.01129
- Liu, Y., Lin, Y., Gao, S., Li, Z., Ma, J., Deng, M., et al. (2017). A genome-wide association study of 23 agronomic traits in Chinese wheat landraces. *Plant J.* 91:861–873. doi: 10.1111/tpj.13614
- Loos, R. J. F. (2020). 15 years of genome-wide association studies and no signs of slowing down. *Nat. Commun.* 11:5900. doi: 10.1038/s41467-020-19653-5
- Mackay, T. F. (2004). The genetic architecture of quantitative traits: lessons from *Drosophila*. *Curr Opin Genet Dev.* 14, 253–257. doi: 10.1016/j.gde.2004.04.003
- Mackay, T. F. C., Stone, E. A., and Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10, 565–577. doi: 10.1038/nrg2612
- Malysheva-Otto, L. V., Ganai, M. W., and Röder, M. S. (2004). Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). *BMC Genet.* 7:6. doi: 10.1186/1471-2156-7-6
- Marshall, A., Aalen, R. B., Audenaert, D., Beeckman, T., Broadley, M. R., Butenko, M. A., et al. (2012). Tackling drought stress: receptor-like kinases present new approaches. *Plant Cell* 24, 2262–2278. doi: 10.1105/tpc.112.096677
- Martinant, J. P., Billot, A., Bouguennec, A., Charmet, G., Saulnier, L., and Branlard, G. (1999). Genetic and environmental variations in water-extractable arabinoxylans content and flour extract viscosity. *J. Cereal Sci.* 30, 45–48. doi: 10.1006/jcrs.1998.0259
- Martis, M. M., Zhou, R., Haseneyer, G., Schmutzer, T., Vrána, J., Kubaláková, M., et al. (2013). Reticulate evolution of the rye genome. *Plant Cell.* 25, 3685–3698. doi: 10.1105/tpc.113.114553
- Melchinger, A. E. (1999). “Genetic diversity and heterosis,” in *The Genetics and Exploitation of Heterosis in Crops*, eds C. G. Coors and S. Pandey (Madison, WI: American Society of Agronomy), 99–118.
- Melz, G., Melz, G., and Hartman, F. (2003). Genetics of a male-sterile rye of ‘G-type’ with results of the first F1-hybrids. *Plant Breed. Seed Sci.* 47, 47–55.
- Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001
- Miedaner, T., Haffke, S., Siekmann, D., Fromme, F. J., Roux, S. R., and Hackauf, B. (2018). Dynamic quantitative trait loci (QTL) for plant height predict biomass yield in hybrid rye (*Secale cereale* L.). *Biomass Bioenergy* 115, 10–18. doi: 10.1016/j.biombioe.2018.04.001
- Miedaner, T., Hertz, C. P., Goßlau, H., Wilde, P., and Hackauf, B. (2017). Correlated effects of exotic pollen-fertility restorer genes on agronomic and quality traits of hybrid rye. *Plant Breed.* 136, 224–229. doi: 10.1111/pbr.12456
- Miedaner, T., Hübner, M., Korzun, V., Schmiedchen, B., Bauer, E., Haseneyer, G., et al. (2012). Genetic architecture of complex agronomic traits examined in two testcross populations of rye (*Secale cereale* L.). *BMC Genomics* 13:706. doi: 10.1186/1471-2164-13-706
- Miedaner, T., Mirdita, V., Rodemann, B., Drobeck, T., and Rentel, D. (2010). Genetic variation of winter rye cultivars for their ergot (*Claviceps purpurea*) reaction tested in a field design with minimized interplot interference. *Plant Breed.* 129, 58–62. doi: 10.1111/j.1439-0523.2009.01646.x
- Miedaner, T., Schwegler, D. D., Wilde, P., and Reif, J. (2014). Association between line *per se* and testcross performance for eight agronomic and quality traits in winter rye. *Theor. Appl. Gen.* 127, 33–41. doi: 10.1007/s00122-013-2198-2
- Milczarski, P., Hanek, M., Tyrka, M., and Stojalowski, S. (2016). The application of GBS markers for extending the dense genetic map of rye (*Secale cereale* L.) and the localization of the *Rfc1* gene restoring male fertility in plants with the C source of sterility-inducing cytoplasm. *J. Appl. Genet.* 57, 439–451. doi: 10.1007/s13353-016-0347-4
- Muqaddasi, Q. H., Brassac, J., Ebmeyer, E., Kollers, S., Korzun, V., Argillier, O., et al. (2020). Prospects of GWAS and predictive breeding for European winter wheat's grain protein content, grain starch content, and grain hardness. *Sci. Rep.* 10:12541. doi: 10.1038/s41598-020-69381-5
- Nadolska-Orczyk, A., Rajchel, I. K., Orczyk, W., and Gasparis, S. (2017). Major genes determining yield-related traits in wheat and barley. *Theor. Appl. Genet.* 130, 1081–1098. doi: 10.1007/s00122-017-2880-x
- Nei, M. (1987). *Molecular Evolutionary Genetics*. New York, NY: Columbia University Press.
- Oest, M., Bindrich, U., Voß, A., Kaiser, H., and Rohn, S. (2020). Rye bread defects: analysis of composition and further influence factors as determinants of dry-baking. *Foods* 9:1900. doi: 10.3390/foods9121900
- Orr, H. A. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52, 935–949. doi: 10.1111/j.1558-5646.1998.tb01823.x
- Papouov, I. A., Lebedinskai, S., and Koshkin, E. I. (1999). Growth analysis of solution culture-grown winter rye, wheat and Triticale at different relative rates of nitrogen supply. *Ann. Bot.* 84, 467–473. doi: 10.1006/anbo.1999.0935
- Pasam, R. K., Sharma, R., Malosetti, M., van Eeuwijk, F. A., Haseneyer, G., Kilian, B., et al. (2012). Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biol.* 12:16. doi: 10.1186/1471-2229-12-16
- Pont, C., Wagner, S., Kremer, A., Orlando, L., Plomion, C., and Salse, J. (2019). Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. *Genome Biol.* 20:29. doi: 10.1186/s13059-019-1627-1
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945
- R Core Team (2017). *R: A Language and Environment for Statistical Computing, v.3.4.1 Edn*. Vienna: R Foundation for Statistical Computing. Available online at: <https://www.r-project.org/>
- Rabanus-Wallace, M. T., Hackauf, B., Mascher, M., Lux, T., Wicker, T., Gundlach, H., et al. (2021). Chromosome-scale genome assembly provides insights into rye biology, evolution, and agronomic potential. *Nat. Genet.* 53, 564–573. doi: 10.1038/s41588-021-00807-0
- Rakoczy-Trojanowska, M., Krajewski, P., Bocianowski, J., Schollenberger, M., Wakuliński, W., Milczarski, P., et al. (2017). Identification of single nucleotide polymorphisms associated with brown rust resistance, α -amylase activity and pre-harvest sprouting in rye (*Secale cereale* L.). *Plant Mol. Biol. Rep.* 35, 366–378. doi: 10.1007/s11105-017-1030-6
- Rakszegi, M., Lovegrove, A., Balla, K., Láng, L., Bedo, Z., Veisz, O., et al. (2014). Effect of heat and drought stress on the structure and composition of arabinoxylan and β -glucan in wheat grain. *Carbohydr. Polym.* 102, 557–565. doi: 10.1016/j.carbpol.2013.12.005
- Remington, D. L., and Purugganan, M. D. (2003). Candidate genes, quantitative trait loci, and functional trait evolution in plants. *Int. J. Plant Sci.* 164(3 Suppl.), S7–S20. doi: 10.1086/367812
- Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., et al. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11479–11484. doi: 10.1073/pnas.201394398
- Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., et al. (2020). Breeder friendly phenotyping. *Plant Sci.* 295:110396. doi: 10.1016/j.plantsci.2019.110396
- Rice, B., and Lipka, A. E. (2019). Evaluation of RR-BLUP genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. *Plant Genome* 12, 1–14. doi: 10.3835/plantgenome2018.07.0052
- Rieseberg, L. H., and Blackman, B. K. (2010). Speciation genes in plants. *Ann. Bot.* 106, 439–455. doi: 10.1093/aob/mcq126

- Rode, J., Ahlemeyer, J., Friedt, W., and Ordon, F. (2012). Identification of marker-trait associations in the German winter barley breeding gene pool (*Hordeum vulgare* L.). *Mol. Breed.* 30, 831–843. doi: 10.1007/s11032-011-9667-6
- Roff, D. A. (2007). A centennial celebration for quantitative genetics. *Evolution* 61, 1017–1032. doi: 10.1111/j.1558-5646.2007.00100.x
- Rosicka-Kaczmarek, J., Komisararczyk, A., Nebesny, E., and Makowski, B. (2016). The influence of arabinoxylans on the quality of grain industry products. *Eur. Food Res. Technol.* 242, 295–303. doi: 10.1007/s00217-015-2549-0
- Sansaloni, C., Franco, J., Santos, B., Percival-Alwyn, L., Singh, S., Petroli, C., et al. (2020). Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. *Nat. Commun.* 11:4572. doi: 10.1038/s41467-020-18404-w
- Schils, R., Olesen, J. E., Kersebaum, K.-C., Rijka, B., Oberforster, M., Kalyada, V., et al. (2018). Cereal yield gaps across Europe. *Eur. J. Agron.* 101, 109–120. doi: 10.1016/j.eja.2018.09.003
- Schittenhelm, S., Kraft, M., and Wittich, K. P. (2014). Performance of winter cereals grown on field-stored soil moisture only. *Eur. J. Agron.* 52, 247–258. doi: 10.1016/j.eja.2013.08.010
- Schreiber, M., Himmelbach, A., Börner, A., and Mascher, M. (2018). Genetic diversity and relationship between domesticated rye and its wild relatives as revealed through genotyping-by-sequencing. *Evol. Appl.* 12, 66–77. doi: 10.1111/eva.12624
- Semagn, K., Babu, R., and Hearne, S. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Over view of the technology and its application in crop improvement. *Mol. Breed.* 33, 1–14. doi: 10.1007/s11032-013-9917-x
- Shaaf, S., Bretani, G., Biswas, A., Fontana, I. M., and Rossini, L. (2019). Genetics of barley tiller and leaf development. *J. Integr. Plant Biol.* 61, 226–256. doi: 10.1111/jipb.12757
- Sidhu, J. S., Ramakrishnan, S. M., Ali, S., Bernardo, A., Bai, G., Abdullah, S., et al. (2019). Assessing the genetic diversity and characterizing genomic regions conferring Tan Spot resistance in cultivated rye. *PLoS ONE* 14:e0214519. doi: 10.1371/journal.pone.0214519
- Skoryk, V. V., Skoryk, V. V., Simonenko, N. V., and Skoryk, O. P. (2010). Genetics characteristics of the donor for dominant short stem and large grain winter rye (*Secale cereale* L.). *Plant Var. Stud. Prot. J. Appl. Res.* 1, 5–12. doi: 10.21498/2518-1017.1(1)0.2010.59362
- Song, S., Tian, D., Zhang, Z., Hu, S., and Yu, J. (2018). Rice genomics: over the past two decades and into the future. *Genomics Proteomics Bioinformatics* 16, 397–404. doi: 10.1016/j.gpb.2019.01.001
- Spindel, J., Begum, H., Akdemir, D., Collard, B., Redoña, E., and Jannink, J.-L. (2016). Genome-wide prediction models that incorporate *de novo* GWAS are a powerful new tool for tropical rice improvement. *Heredity* 116, 395–408. doi: 10.1038/hdy.2015.113
- Sprague, G. F., and Tatum, A. L. (1942). General vs. specific combining ability in single crosses of corn. *J. Amer. Soc. Agron.* 34, 923–932. doi: 10.2134/agronj1942.00021962003400100008x
- Stojalowski, S., Kociuba, M., Stochmal, B., Kondziola, M., and Jaciubek, M. (2008). Determining the plasmotypic structure of rye populations by SCAR markers. *J. Appl. Genet.* 49, 229–232. doi: 10.1007/BF03195618
- Targowska-Karasek, M., Bolibok-Bragoszewska, H., and Rakoczy-Trojanowska, M. (2017). DArTseq genotyping reveals high genetic diversity of polish rye inbred lines. *Crop Sci.* 57, 1906–1915. doi: 10.2135/cropsci2016.09.0771
- Tenhaken, R. (2015). Cell wall remodeling under abiotic stress. *Front. Plant Sci.* 5:771. doi: 10.3389/fpls.2014.00771
- Tilley, M. S., Heiniger, R. W., and Crozier, C. R. (2019). Tiller initiation and its effects on yield and yield components in winter wheat. *Agron. J.* 111, 1323–1332. doi: 10.2134/agronj2018.07.0469
- Van Ooijen, J. W. (2006). *JoinMap® 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Wageningen: Kyazma B.V.
- Van Os, H., Stam, P., Visser, R. G., and Van Eck, H. J. (2005). RECORD: a novel method for ordering loci on a genetic linkage map. *Theor. Appl. Genet.* 112, 30–40. doi: 10.1007/s00122-005-0097-x
- Vendelbo, N. M., Sarup, P., Orabi, J., Kristensen, P. S., and Jahoor, A. (2020). Genetic structure of a germplasm for hybrid breeding in rye (*Secale cereale* L.). *PLoS ONE* 15:e0239541. doi: 10.1371/journal.pone.0239541
- Voorrips, R. E. (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77
- Wan, Y., Wang, Y., Shi, Z., Rentsch, D., Ward, J. L., Hassall, K., et al. (2021). Wheat amino acid transporters highly expressed in grain cells regulate amino acid accumulation in grain. *PLoS ONE* 16:e0246763. doi: 10.1371/journal.pone.0246763
- Wang, A., Jiang, Y., Shu, X., Zha, Z., Yin, D., Liu, Y., et al. (2021). Genome-wide association study-based identification genes influencing agronomic traits in rice (*Oryza sativa* L.). *Genomics*, 113, 1396–1406. doi: 10.1016/j.ygeno.2021.03.016
- Wang, B., and Li, J. (2019). Understanding the molecular bases of agronomic trait improvement in rice. *Plant Cell* 31:1416–1417. doi: 10.1105/tpc.19.00343
- Wang, W., Li, G., Zhao, J., Chu, H., Lin, W., Zhang, D., et al. (2014). Dwarf Tiller1, a Wuschel-related homeobox transcription factor, is required for tiller growth in rice. *PLoS Genet.* 10:e1004154. doi: 10.1371/journal.pgen.1004154
- Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49. doi: 10.1038/s41586-018-0063-9
- Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.1111/j.1558-5646.1984.tb05657.x
- White, M. R., Mikel, M. A., de Leon, N., and Kaeppler, S. M. (2020). Diversity and heterotic patterns in North American proprietary dent maize germplasm. *Crop Sci.* 60, 100–114. doi: 10.1002/csc2.20050
- Whitford, R., Fleury, D., Reif, J. C., Garcia, M., Okada, T., Korzun, V., et al. (2013). Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. *J. Exp. Bot.* 64, 5411–5428. doi: 10.1093/jxb/ert333
- Wilke, V., Grone, R., von Felde, A., Abd El-Wahab, A., Wolf, P., and Kamphues, J. (2021). Effects of increasing dietary rye levels on physicochemical characteristics of digesta and its impact on stomach emptying as well as the formation of 'doughballs' in stomachs of young pigs. *J. Anim. Physiol. Anim. Nutr.* doi: 10.1111/jpn.13549
- Wolski, T., Brykczynski, J., and dan Tymieniecka, E. (1972). Heritability of some characters of rye under open polli-nation. *Theor. Appl. Genet.* 42, 168–173. doi: 10.1007/BF00280793
- Wricke, G. (2002). Two major genes for kernel weight in rye. *Plant Breed.* 121:26–28. doi: 10.1046/j.1439-0523.2002.00666.x
- Wu, D., Guo, Z., Ye, J., Feng, H., Liu, J., Chen, G., et al. (2019). Combining high-throughput micro-CT-RGB phenotyping and genome-wide association study to dissect the genetic architecture of tiller growth in rice. *J. Exp. Bot.* 70, 545–561. doi: 10.1093/jxb/ery373
- Xing, Y., and Zhang, Q. (2010). Genetic and molecular bases of rice yield. *Annu. Rev. Plant Biol.* 61, 421–442. doi: 10.1146/annurev-arplant-042809-112209
- Yamamoto, E., Yonemaru, J. I., Yamamoto, T., and Yano, M. (2012). OGRO: the overview of functionally characterized genes in rice online database. *Rice* 5:26. doi: 10.1186/1939-8433-5-26
- Yang, Y., Amo, A., Wei, D., Chai, Y., Zheng, J., Qiao, P., et al. (2021). Large-scale integration of meta-QTL and genome-wide association study discovers the genomic regions and candidate genes for yield and yield-related traits in bread wheat. *Theor. Appl. Genet.* 134, 3083–3109. doi: 10.1007/s00122-021-03881-4
- Yang, Y., Chai, Y., Zhang, X., Lu, S., Zhao, Z., Wei, D., et al. (2020). Multi-locus GWAS of quality traits in bread wheat: mining more candidate genes and possible regulatory network. *Front. Plant Sci.* 11:1091. doi: 10.3389/fpls.2020.01091
- Yao, W., Li, G., Yu, Y., and Ouyang, Y. (2018). funRiceGenes dataset for comprehensive understanding and application of rice functional genes. *GigaScience* 7:gix119. doi: 10.1093/gigascience/gix119
- Yeo, I. Y., Lee, S., Sadeghi, A. M., Beeson, P. C., Hively, W. D., McCarty, G. W., et al. (2014). Assessing winter cover crop nutrient uptake efficiency using a water quality simulation model. *Hydrol. Earth Syst. Sci.* 18, 5239–5253. doi: 10.5194/hess-18-5239-2014
- Yonemaru, J. I., Yamamoto, T., Fukuoka, S., Uga, Y., Hori, K., and Yano, M. (2010). Q-TARO: QTL annotation rice online database. *Rice* 3, 194–203. doi: 10.1007/s12284-010-9041-z
- Zheng, M., and Zhuo, M. (2017). Analysis of corrections methods in genome-wide association studies. *Adv. Comp. Sci. Res.* 76, 439–442. doi: 10.2991/emim-17.2017.88

Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. *Cell* 167, 313–324. doi: 10.1016/j.cell.2016.08.029

Conflict of Interest: DS and FF are employed by HYBRO Saatzucht GmbH & Co. KG, and AK is employed by Diversity Arrays Technologies. The companies have a commercial interest in the results for application in variety development and for the provision of molecular genetic services. This does not alter the adherence of the authors to all Frontiers in Plant Science policies on sharing data and materials.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Siekmann, Jansen, Zaar, Kilian, Fromme and Hackauf. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genome-Wide Identification of Wheat ZIP Gene Family and Functional Characterization of the *TaZIP13-B* in Plants

Song Li, Zihui Liu, Linlin Guo, Hongjie Li, Xiaojun Nie, Shoucheng Chai and Weijun Zheng*

State Key Laboratory of Crop Stress Biology for Arid Areas, College of Agronomy, Northwest A&F University, Yangling, China

OPEN ACCESS

Edited by:

Shuyu Liu,
Texas A&M AgriLife Research,
The Texas A&M University System,
United States

Reviewed by:

Kaushal Kumar Bhati,
Catholic University of Louvain,
Belgium
Ruibo Hu,
Qingdao Institute of Bioenergy and
Bioprocess Technology, Chinese
Academy of Sciences (CAS), China

*Correspondence:

Weijun Zheng
zhengweijun@nwfau.edu.cn

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 27 July 2021

Accepted: 11 October 2021

Published: 03 November 2021

Citation:

Li S, Liu Z, Guo L, Li H, Nie X,
Chai S and Zheng W (2021)
Genome-Wide Identification of
Wheat ZIP Gene Family and
Functional Characterization of the
TaZIP13-B in Plants.
Front. Plant Sci. 12:748146.
doi: 10.3389/fpls.2021.748146

The ZIP (Zn-regulated, iron-regulated transporter-like protein) transporter plays an important role in regulating the uptake, transport, and accumulation of microelements in plants. Although some studies have identified ZIP genes in wheat, the significance of this family is not well understood, particularly its involvement under Fe and Zn stresses. In this study, we comprehensively characterized the wheat ZIP family at the genomic level and performed functional verification of three *TaZIP* genes by yeast complementary analysis and of *TaZIP13-B* by transgenic *Arabidopsis*. Totally, 58 *TaZIP* genes were identified based on the genome-wide search against the latest wheat reference (IWGSC_V1.1). They were then classified into three groups, based on phylogenetic analysis, and the members within the same group shared the similar exon-intron structures and conserved motif compositions. Expression pattern analysis revealed that the most of *TaZIP* genes were highly expressed in the roots, and nine *TaZIP* genes displayed high expression at grain filling stage. When exposed to ZnSO_4 and FeCl_3 solutions, the *TaZIP* genes showed differential expression patterns. Additionally, six ZIP genes responded to zinc-iron deficiency. A total of 57 miRNA-*TaZIP* interactions were constructed based on the target relationship, and three miRNAs were downregulated when exposed to the ZnSO_4 and FeCl_3 stresses. Yeast complementation analysis proved that *TaZIP14-B*, *TaZIP13-B*, and *TaIRT2-A* could transport Zn and Fe. Finally, overexpression of *TaZIP13-B* in *Arabidopsis* showed that the transgenic plants displayed better tolerance to Fe/Zn stresses and could enrich more metallic elements in their seeds than wild-type *Arabidopsis*. This study systematically analyzed the genomic organization, gene structure, expression profiles, regulatory network, and the biological function of the ZIP family in wheat, providing better understanding of the regulatory roles of *TaZIPs* and contributing to improve nutrient quality in wheat crops.

Keywords: wheat, ZIP gene family, expression profiles, yeast complementation, Zn/Fe stress, transgenic *Arabidopsis*, micro elements

INTRODUCTION

Zinc (Zn) and iron (Fe), both essential in biochemical activities, are required for plant growth and development. Zn is an essential component for the metabolic enzymes that regulate enzymatic activity (Maret, 2004; Welch and Graham, 2004). Iron is also important for many enzymes, including cytochrome oxidase, peroxidase, and catalase, all of which play an important role in respiratory electron transport (Sadeghzadeh, 2013; Pinson et al., 2015). During photosynthesis, Zn is linked with carbohydrate inversion and is involved in chlorophyll synthesis. Fe is an essential for some chlorophyll protein complexes in chloroplasts (Palmgren et al., 2008). Although plant growth and development requires Zn and Fe, excessive amounts of Zn and Fe are harmful to the plant's biological processes (Briat and Lebrun, 1999). As a result, plant cells have evolved multifunctional transport networks to balance the absorption, utilization, and storage of these metal trace elements (Kambe et al., 2004; Taylor et al., 2004). These systems include the ZIP (Zn-regulated, iron-regulated transporter-like protein), CDF (Cation-Diffusion Facilitator), and HMA (Heavy Metal ATPase) proteins (Colangelo and Gueriot, 2006; Ajeesh Krishna et al., 2020).

In plant, ZIP transporters are involved in transporting iron and metallic ions. Most ZIPs are composed of 326–425 amino acid residues. The number of transmembrane domains (TM) in ZIP transporters ranged from 7 to 8, while the TM length between III and IV varies and contains multiple histidine residues (Gueriot, 2000), that ZIP transporters combine metal ions to form octahedral, tetrahedral, and plane structures (Kavitha et al., 2015). The first ZIP gene was reported in *Arabidopsis*, and many ZIP genes have been identified in recent years (Eide et al., 1996; Gueriot, 2000; Pence et al., 2000; Tiong et al., 2015). The proteins of these genes can transport various divalent cations, including Fe^{2+} , Zn^{2+} , Mn^{2+} , and Cd^{2+} . Sixteen ZIP genes have been found in rice and *Arabidopsis* (Mäser et al., 2001; Chen et al., 2008).

The first ZIP gene identified was *AtIRT1* in *Arabidopsis*, which primarily transports iron and is expressed in the roots. *AtIRT1* was upregulated under iron-deficient conditions and is upregulated when exposed to a nickel solution (Eide et al., 1996). *AtIRT1* was proven to transport Fe and Zn by yeast complementation assays. Further research demonstrated the *irt1* mutant leaves were severely etiolate, with the leaf iron content decreased by 70% compared to the wild type (WT; Vert et al., 2002). As the *AtIRT1* transformed into *irt1* mutant, the etiolation phenotype was alleviated (Krämer et al., 2007). *AtIRT2* is another IRT gene in *Arabidopsis* that has a similar function to *AtIRT1*, it can restore the ability to transport iron in yeast mutants (Vert et al., 2001). *AtZIP1* and *AtZIP2* are two genes that primarily transport Zn. *AtZIP1* is primarily expressed in the root and leaf vein, while *AtZIP2* is highly expressed in the root column (Milner et al., 2013). Subcellular localization analysis revealed that the protein of *AtZIP1* is located on the vacuole membrane and the protein of *AtZIP2* is located on the plasma membrane. This difference in protein localization implies that *AtZIP1* and *AtZIP2* function keep differently. Functional validation revealed that *AtZIP1* plays a

key role in the reactivation of metal ions transported from the vacuoles to the root cytoplasm, whereas *AtZIP2* is involved in Mn and Zn absorption from the roots. Both genes are crucial for a plant to absorb Mn and Zn through its root and transport them from the roots to the leaves (Milner et al., 2013). Previous studies have demonstrated that certain ZIP genes are involved in the response to Zn-deficiency in *Arabidopsis* (Grotz and Gueriot, 2006; Lee et al., 2010b).

OsIRT1 and *OsIRT2* are the primary transporters of Fe in rice (Ishimaru et al., 2006; Li et al., 2019a). These two genes are mainly expressed in the roots and are significantly upregulated when rice is exposed to Fe-deficient conditions (Ishimaru et al., 2006; Itai et al., 2013). Overexpression of *OsIRT1* in rice causes rice sensitivity to Zn and Cd, while also increase resistance to Fe-deficient stress (Ishimaru et al., 2007; Nakanishi et al., 2010). At the seedling stage, there is no significant difference between the phenotype of the overexpression and the wild-type varieties; however, at the adult stage, the overexpression variety had shorter and fewer tillers and lower yields compared to the wild type, while the Fe and Zn contents in the grains increased (Lee and An, 2009; Lee et al., 2010b). *OsIRT2* was similar to *OsIRT1* in that the transport capability of Mn is less than *OsIRT1* (Nakanishi et al., 2010). Previous studies have revealed that the family members *OsZIP3*, *OsZIP4*, *OsZIP5*, and *OsZIP8* also transported Zn in rice (Ramesh et al., 2003; Lee and An, 2009; Lee et al., 2010a,b; Kavitha et al., 2015; Sasaki et al., 2015). Nine ZIP genes were identified in maize and were located on the plasma membrane and endomembrane system. Yeast complementation demonstrated that all ZmZIP proteins can restore the iron transporter mutant *fet3fet4*, and that *ZmIRT1* showed the strongest propagation under both Zn- and Fe-limited conditions (Li et al., 2013).

ZIP proteins have been widely investigated in model plants such as rice, maize, and *Arabidopsis* (Eide et al., 1996; Vert et al., 2001, 2002; Ishimaru et al., 2006; Li et al., 2013). However, few ZIP genes have been reported in wheat plants.

Sixteen ZIP genes have been identified in the wheat genome, though few studies have performed the expression analysis on these genomes (Tiong et al., 2015; Evens et al., 2017). Five ZIP genes in wheat were demonstrated to transport zinc and iron, these five ZIP genes were selected for analysis using yeast complementation since their sequence was similar to Zn-transporting ZIPs from *Arabidopsis*, rice, and barley (Tiong et al., 2014, 2015; Evens et al., 2017).

Wheat is a worldwide staple crop, feeding approximately 35% of the world's population (Peng et al., 2011). As breeding technology and cultivation programs have increased, the production of wheat has risen. However, its nutritional quality has not improved: the Fe and Zn contents in wheat cannot meet human needs. Approximately two billion people suffer from Zn and Fe deficiency in South Asia and Sub-Saharan Africa. This deficiency has been called “hidden hunger,” and makes induces weight loss, cognitive impairment, anti-spasmodic decline, and often occurs in pregnant women, infants, and adolescents (Morgounov et al., 2007). Genetic engineering is the most convenient, effective, and durable method of increasing the Zn and Fe content in wheat grain. Therefore, it is critical

to identify the genes involved in the uptake, transport, and enrichment of Zn and Fe in wheat. In this study, the ZIP genes in wheat were analyzed at the genomic level, three TaZIP genes were functionally validated by yeast complementation, and *TaZIP13-B* function was further verified by transgenic *Arabidopsis*. This research aims to uncover new candidate genes to improve the nutritional quality of wheat.

MATERIALS AND METHODS

Plant Materials

Two wheat varieties (ZhongMai175 and Xiaobaimai) and one rice variety (*Oryza sativa* L. *japonica*. cv. Nipponbare) were utilized in this research. ZhongMai175 is a high Zn wheat variety that allowing for analysis of the expression of TaZIP genes in wheat (He et al., 2015). This line was planted at the experimental station of Northwest A&F University, Yangling, China (34°20'N, 108°24'E). Each row was 1 m wide with four duplicates, while the row spacing was 0.25 m and the plant spacing in each row was 0.05 m. At the 7 days after flowering stage (7 DAF), we collected grain sample from one plant in each row which was mixed with liquid nitrogen. One week later, the second sample was obtained (14 DAF). Four samples were collected in this study; there are 7, 14, 21, and 28 DAF.

Xiaobaimai is a landrace, drought-tolerant wheat variety found in PingYao city (ShanXi Province, China). This variety contained low zinc and iron in grain. Culturing in a glass petri dish with two layers of filter paper containing 1/4 Hoagland solution and place at climate chambers (RXZ-500D-LED, Ning Bo) at a light/dark cycle of 16/8 h at 24°C for 10 days (Seedling two leaf). Then for analyzed the expression under Fe and Zn stress, wheat treated with Hoagland medium containing different concentrations of ZnSO₄ and FeCl₃ solutions for 1 h. In this study, the concentration of ZnSO₄ and FeCl₃ solutions was 0.05, 0.5, 50 μmol/L. For investigated the expression under Zn- and Fe- deficient conditions, we treated it with Hoagland medium lacking ZnSO₄ (Zn-deficient), Fe (III)-EDTA (Fe-deficient).

Identification and Bioinformatics Analyses of TaZIP Genes

The sequence of the wheat proteins was downloaded from the Ensembl Plant database,¹ after which the HMM profile for the ZIP DNA-binding domain (PF02535) was downloaded from the Pfam v31.0 database² to search against the plant protein sequences using a threshold of $E < 1e^{-5}$ (Finn et al., 2016). Blast and manual corrections were then performed to remove alternative events and redundancy. The OsZIP genes were downloaded from the NCBI database,³ according to the methods used by Chen and Tiong (Chen et al., 2008; Tiong et al., 2014). The NJ phylogenetic tree was constructed with MEGA 7 and

EvolView⁴ based on the wheat and rice protein sequences, with 1,000 bootstrap replicates. Putative TMHs of TaZIPs were predicted using the TMHMM Server v2.0 (Krogh et al., 2001). Subcellular location was predicted by the WoLF PSORT.⁵

Gene Structure and Conserved Motif Analyses

Gene structure was analyzed by GSDS.⁶ Protein conserved motifs were predicted using the MEME Suite web server,⁷ with the number of motifs set to 10, at a width range from 5 to 200 amino acids.

TaZIP Genes and miRNA Co-expression Networks Construction

The miRNA target to the TaZIP genes was searched using the psRNATarget tool (Dai and Zhao, 2011), and the TaZIP cascade transcript was submitted in the miRBase. The cytoscape tool⁸ was used to visualize the regulatory network of the-miRNA and TaZIP genes.

RNA and miRNA Isolate and Expression Pattern Analysis

For wheat tissue expression pattern analysis, the expression pattern data were downloaded from the RNA-seq database.⁹ For the expression pattern at grain filling stage and under Zn and Fe stress assay, RT-qPCR was used. Total RNA was isolated from the wheat grains and wheat, rice leaves using an RNAPrep Pure Plant Kit (Tiangen, Beijing, China) and from the seedlings with TRIZOL (Takara, Dalian, China). cDNA synthesis was performed in a 20 μl reaction mixture containing 1 μg of total RNA and a mixture of TIANscript RT Kit (Tiangen, Beijing, China). The real-time PCR mixture contained 1 μl cDNA, 1 μl forward and reverse primers, and 17 μl SYBR Green (Tiangen, Beijing, China). Real-time qPCR was performed in an ABI7300 (Thermo Fisher Scientific, United States) Real-Time Thermal Cycler and repeated three times. The actin of the wheat genes (Gene ID: AB181991) was used as a control. The 2^{-ΔΔCt} method was used for fluorescence quantitative data analysis (Livak and Schmittgen, 2001).

The miRNAs of the wheat seedling under ZnSO₄ and FeCl₃ stresses were extracted using a miRcute miRNA Isolation Kit (Tiangen KR211, Beijing, China). miRNA-cDNA synthesis was performed with miRcute and miRNA First-Strand cDNA Kit (Tiangen KR211, Beijing, China). The real-time reaction mixture was performed with miRcute Plus miRNA qPCR Kit (SYBR Green; Tiangen FP411, Beijing, China) with three biological replicates. The qPCR reaction conditions were 95°C for 15 min, followed by 45 cycles of 94°C for 20 s, and 58–60°C for 34 s. Data of miRNA-qPCR were analyzed using the 2^{-ΔΔCt} method.

⁴<https://www.evolgenius.info/evolview/#login>

⁵<https://wolfsort.hgc.jp/>

⁶<http://gsds.cbi.pku.edu.cn/>

⁷<http://meme-suite.org/>

⁸<http://www.cytoscape.org/>

⁹<http://www.wheat-expression.com/>

¹<http://plants.ensembl.org/index.html>

²<http://pfam.xfam.org/>

³<https://www.ncbi.nlm.nih.gov/>

Cloning of TaZIP and OsZIP Genes

The CDS and ORF sequences were obtained from the Wheat Sequence database.¹⁰ The primers were designed with Oligo 7, with three genes cloned. In this step, the RNA of Xiaobaimai and *Oryza sativa* L. seedling was used for cDNA synthesis. The PCR reaction solution was 50 µl and contained the following: 5 µl cDNA as an amplification template, 2.5 µl forward and reverse primers, 25 µl 2× master Mix, and 15 µl Nuclease free water (NEB, United States). The reaction solution was performed on a DNA amplification machine (Thermo Fisher Scientific, United States). PCR amplification procedures were as follows: initial denaturation at 98°C for 30 s, followed by 35 cycles of denaturation at 98°C for 10 s, annealing at 60°C for 20 s, extension at 72°C for 30 s, while the final extension was at 72°C for 2 min. After amplification, we added 7 µl purple 2-Log Ladder (NEB, United States) to the PCR products and separated them on 1.5% agarose gel for 30 min at 120 V. After they were separated, the PCR products were purified using a Universal DNA Purification Kit (Tiangen, Beijing, China) and connected to the cloning vector pLB (Tiangen, Beijing, China), and sequenced.

Yeast Complementation

Specific primers were designed for PCR amplification and constructing the expression vector. The PCR procedure is the same as above, except for the annealing, which took place at 70°C for 20 s. The PCR products were spread on agarose gel and linked to the *Bam*H I site of the yeast expression vector pDRI95 (PLASMID, China). They were subsequently sequenced and transformed into yeast competent cells. The yeast competent cells were prepared according to the methods used by Gietz and Schiestl (1995). Three yeast strains were used in this experiment: DY1455 (MATa ade6 can1 his3 leu2 trp1 ura3), *fet3fet4* DEY1453 (MATa/MATa ade2/+ can1/can1 his3/his3 leu2/leu2 trp1/trp1 ura3/ura3 *fet3-2::HIS3/fet3-2::HIS3 fet4-1::LEU2/fet4-1::LEU2*), and *zrt1zrt2* ZHY3 (MATa ade6 can1 his3 leu2 trp1 ura3 *zrt1::LEU2 zrt2::HIS3*; Li et al., 2013). The pDRI95-TaZIPs were converted to DEY1453 and ZHY3 with the lithium acetate conversion method used by Gietz and Schiestl (1995). To verify the experiment was performed properly, *OsIRT1*, *OsZIP3*, and *OsZIP5* as well as converted to yeast competent cells as positive controls, the wild-type strain DY1455 harboring pDRI95 was also used as a positive control. The empty vector pDRI95 was used as a negative control converted to two yeast mutants. Transformed cells were coated on the selective SD-URA solid medium without corresponding amino acids. To verify the gene function, we diluted the yeast liquid OD₆₀₀ to 1, 0.1, 0.01, 0.001, and dropped 10 µl yeast liquid onto a different medium. The yeast strain of *zrt1zrt2* ZHY3 was grown on an SD/-ura medium (pH 4.4) supplemented with 0.4 mM EDTA or 300 µM ZnSO₄. The yeast strain of *fet3fet4* DEY1453 was grown on SD/-ura medium (pH 5.8) containing 50 mM 2-(4-Morpholino) ethanesulfonic acid (MES) supplemented or 200 µM FeCl₃.

Phenotype Analysis in *Arabidopsis*

Specific primers were used for vector construction, the expression vector PCAMBIA1302 and the *Nco* I site were used for gene construction. After sequencing, the overexpression plasmid of PCAMBIA1302-TaZIP13 was transferred into the GV3101 (*Agrobacterium tumefaciens*) strain and transformed into *Arabidopsis*. The transgenic lines were cultured in a light temperature incubator until they were propagated for the third generation. The homozygous plants of the T3 progeny and WT were used for further study.

For the germination assays, the seeds of wild type and transgenic lines were surface sterilized and kept at 4°C for 72 h in the dark before germination. About 25 seeds of every genotype were sown on the same plate containing different concentration of FeCl₃ and ZnSO₄ solution MS medium at 22/20°C (day/night) with a photoperiod of 16/8 h (day/night) for 7 days. Each day germinated seeds with protruded radicles were counted. The concentration of FeCl₃ and ZnSO₄ solution was 0, 50, 200, 300 µmol/L. After germinated seeding were counted, four *Arabidopsis* lines were cultured until they were at six leaf stage, then the root length of 30 seedlings from each line was measured and photographed. On the other hand, four lines from MS medium were transplanted into soil and treated with different concentrations of FeCl₃ and ZnSO₄ to analyze the tolerance. In this study, the concentration of ZnSO₄ and FeCl₃ was 200, 300, 400 µmol/L. After 2 weeks, a 0.15 g of each sample leaf was collected and determine the chlorophyll content according to the methods used by Richardson et al. (2010). And three leaves of each line were collected and taken the midsection epidermis of the leaves to observe the stomas.

Seeds of Light Microscopy, and Root Length, Fe and Zn Contents

Four lines were treated with water until maturity. Then seeds from the siliques located in the basis of a major inflorescence were selected for observation. Seeds from WT and transgenic lines were randomly selected, then photographed using stereomicroscope (Olympus mx7) and a test instrument (WAN SHENG, HiCC-A). For the Zn and Fe content assays, 0.15 g plant shoots, roots, and seeds of *Arabidopsis* were collected and digested in 2 ml HNO₃ overnight, then 2 ml H₂O₂ was added and completed the digestion by microwave treatment, after which the digests were diluted with Millipore-purified water and filtered. The volume was then adjusted to 25 ml, was ICP-OES analyses using an ICAP 6000 Series spectrometer (Thermo-Fisher; Hansen et al., 2013). For metal content measurements, three samples were conducted for each line. Each sample (shoots, roots, and seeds, respectively) was a mix of 15 plants.

Statistical Analysis

The length of roots, length and width of seeds were counted by ImageJ software (Rueden et al., 2017). Data were analyzed and graphs were drawn using Excel 2019 (Microsoft Corporation,

¹⁰<https://wheat-urgi.versailles.inra.fr/Tools>

United States). In all graphs, error bars indicate standard deviation, and significant differences are indicated with * $p < 0.05$ or ** $p < 0.01$.

RESULTS

Identification and Classification of ZIP Genes in Wheat

A total of 58 ZIP genes were identified using a whole-genome search (Figure 1; Supplementary Table S1; Supplementary Figures S1, S2), of which 44 wheat ZIP genes were found to share an orthologous relationship with rice. These were named according to rice-related nomenclature (Figure 1). The remaining genes were named from *TaZIP17* to *TaZIP30*, based on their location on the chromosome, from 1A to 7D (Supplementary Figure S2). Our results demonstrated that the *TaZIPs* were

unevenly distributed on the chromosomes, and no ZIP genes located on the fifth chromosome group. The length of the *TaZIP* amino acids ranged from 185 to 577 and contained between 3 and 13 transmembrane domains. Most contained between 7 and 9 TM, while the length between TM-3 and TM-4 varied. Subcellular localization of the *TaZIP* genes was found to be on the plasma membrane and nucleus (Supplementary Table S1).

We constructed the phylogenetic relationship of the wheat ZIPs with rice, maize, and *Arabidopsis* ZIP proteins (Figure 1 and Supplementary Tables S1 and S2). The results revealed that these ZIP proteins were classified into three groups. Group ZIP I and group ZIP III included all species proteins, suggesting that these ZIP proteins have a conserved function in monocots and dicots. Only three monocot species were included in group ZIP II, including *Arabidopsis*, suggesting that these proteins are unique to

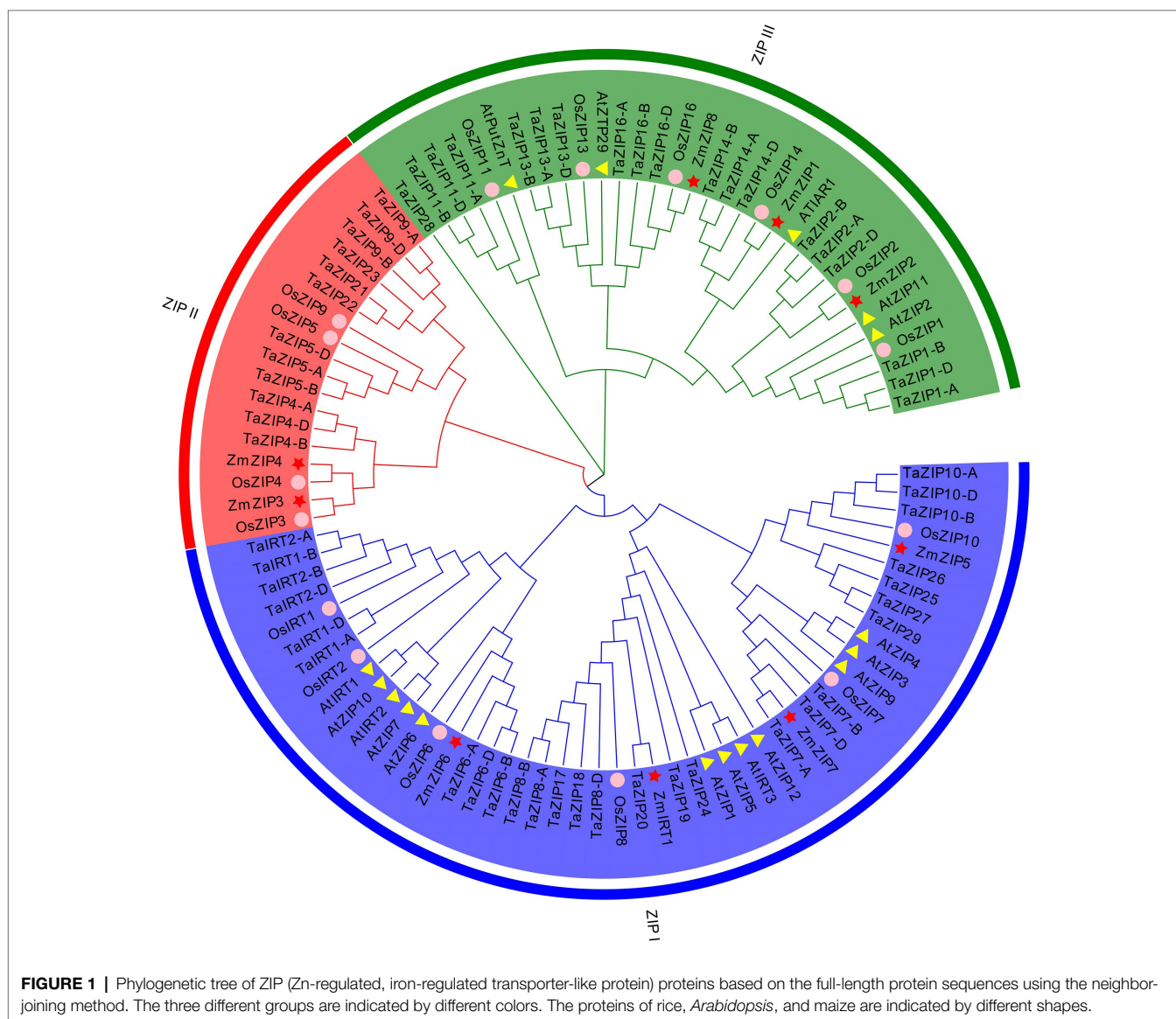
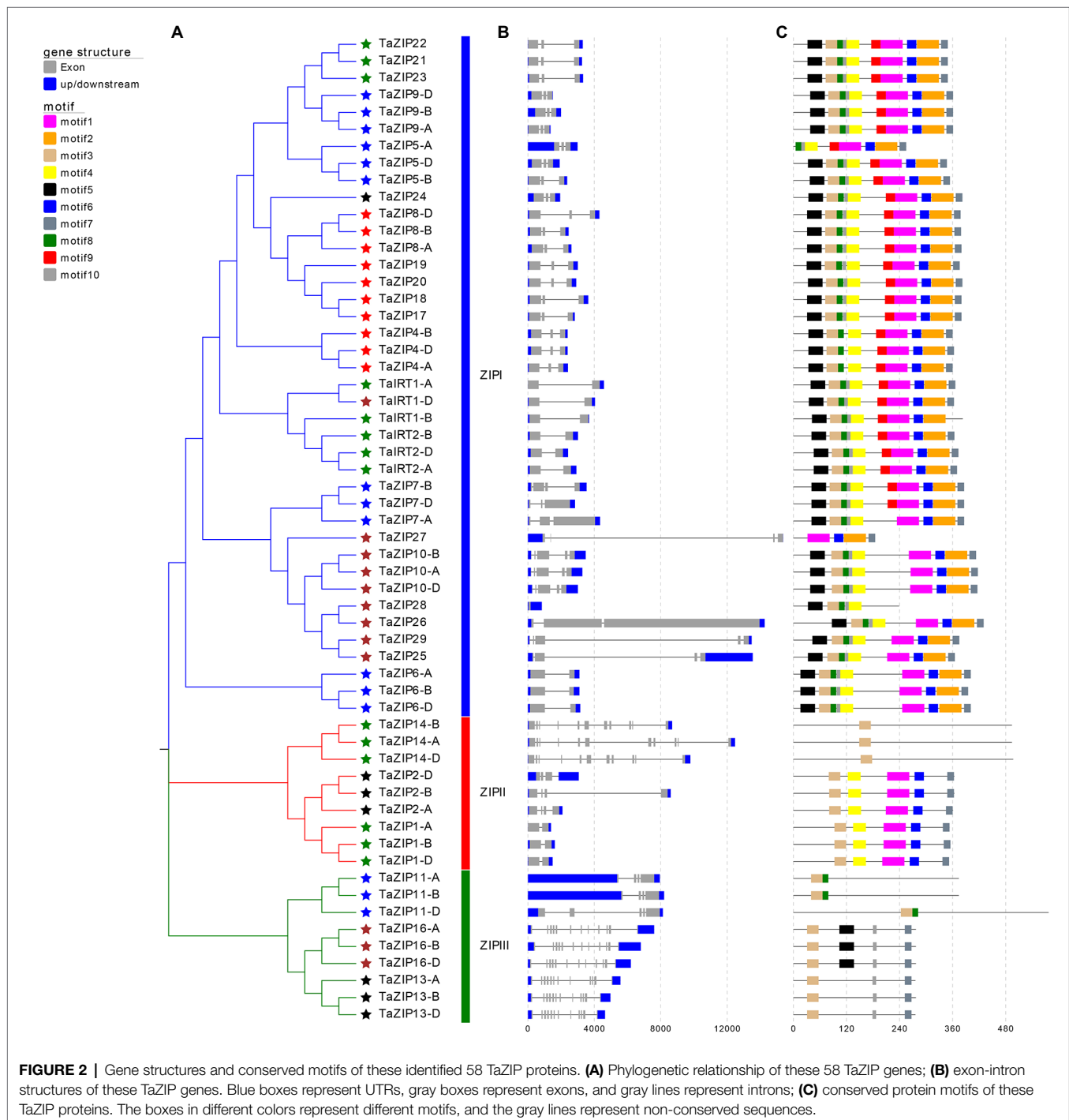


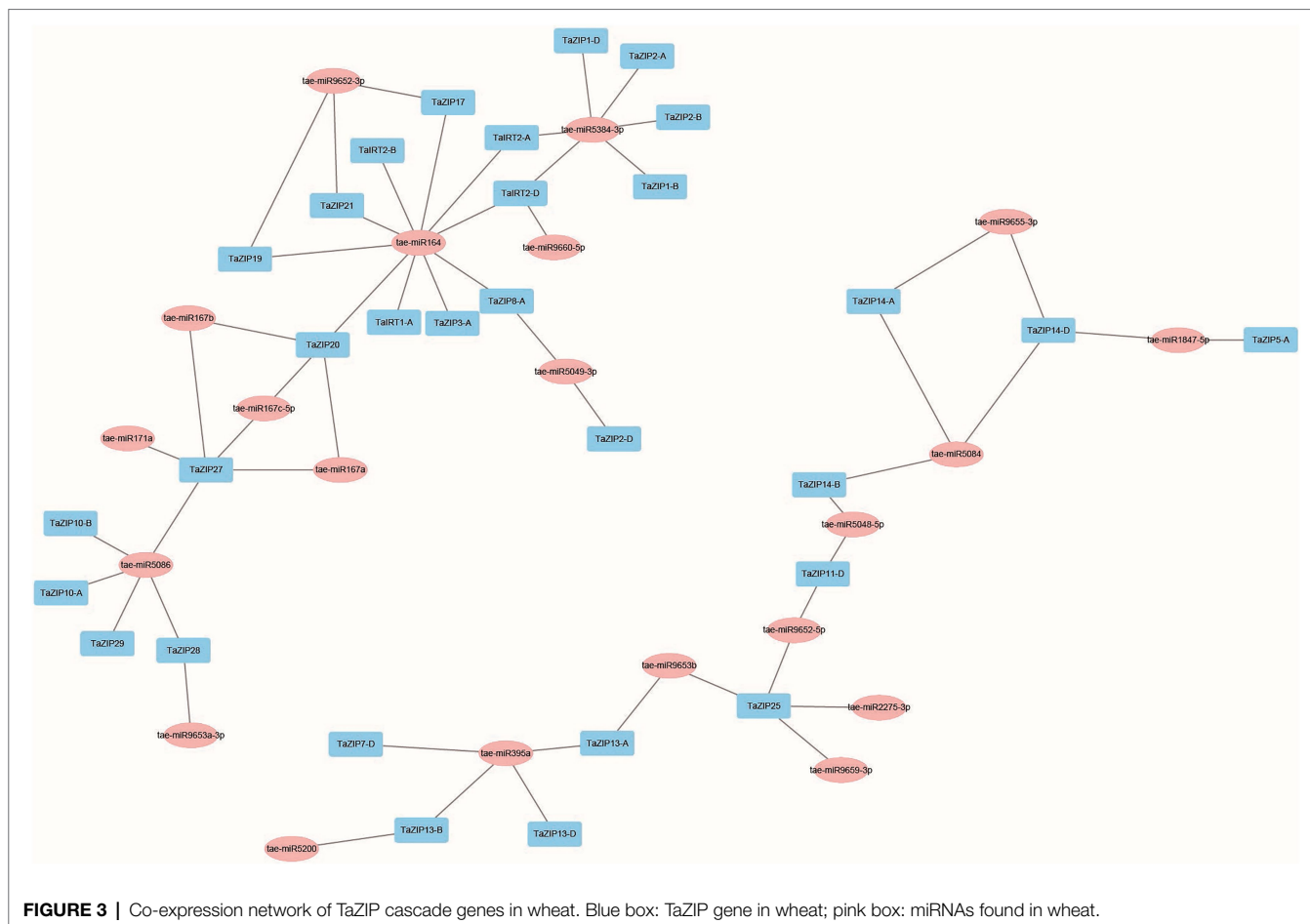
FIGURE 1 | Phylogenetic tree of ZIP (Zn-regulated, iron-regulated transporter-like protein) proteins based on the full-length protein sequences using the neighbor-joining method. The three different groups are indicated by different colors. The proteins of rice, *Arabidopsis*, and maize are indicated by different shapes.

monocot. The wheat proteins are consistent with those of rice and maize, except for *ZmZIP5* and *ZmZIP1*. Six wheat ZIP genes showed a close relationship with *OsIRT1*, *OsIRT2*, *AtIRT1*, and *AtIRT2*, suggesting that these six TaZIPs shared a similar function with *OsIRT1*, *OsIRT2*, *AtIRT1*, and *AtIRT2*, all of which are involved in Fe transport in wheat. The wheat proteins are evenly distributed on three branches.

Gene Structure and Conserved Motifs of TaZIP Genes

To comprehensively understand the function of TaZIPs, we analyzed their gene structure and conserved motifs (**Figure 2**). ZIP size ranged from 836 to 14,494 bp (**Supplementary Table S1**). Of these, the *TaZIP28* gene was the shortest and the *TaZIP27* gene was the longest. The number of introns varied from 0 to 11, and the number of exons ranged from 1 to 12. *TaZIP27*





had the largest size, *TaZIP16-A, B, D* and *TaZIP13-A, B, D* had the most exons (up to 12). In addition, *TaZIP28* has only one exon. Genes sharing a closer phylogenetic relationship had a more similar gene structure.

Using the MEME tool, 10 conserved motifs were identified in the wheat ZIPs (**Figure 2** and **Supplementary Table S3**). The conserved motifs of the same group were similarly organized. Almost all TaZIP proteins possess motif 3 because it contains a histidine residue that binds to metal ions for transmembrane transport. All TaZIP proteins in group ZIPI had motif 1 through motif 7, except for four truncated genes that lacked some of these motifs (*TaZIP5-A*, *TaIRT1-D*, *TaZIP27*, and *TaZIP28*). In group ZIPII, three TaZIP proteins (*TaZIP14-A, B, D*) only had motif 3, while other members have the same motif. In group ZIPIII, motif 8 was detected in *TaZIP11-A, B, D*, and motif 5 was only detected in *TaZIP16-A, B, D*. Group ZIPI had the most motifs, while group ZIPII and ZIPIII had different quantities of motifs.

Network Construction of TaZIP Cascade Genes

The putative miRNA-targeted TaZIP genes were analyzed to assess the network of miRNA and TaZIP genes. Our results demonstrated that 20 miRNAs were predicted to target 30

TaZIP genes, while 28 TaZIP genes were not targeted by miRNA. This could be due to the current limitations on wheat miRNA (**Supplementary Table S4**). Based on the target relationship, 57 miRNA-TaZIP interactions were constructed (**Figure 3**). The wheat ZIP genes were inhibited by miRNA *via* translation (57.89%), while the rest of the genes were inhibited *via* cleavage (42.11%). Additionally, the miRNAs are primarily targeted in the CDS region but ahead of the ZIP domain of the TaZIP genes, silencing gene expression.

Furthermore, we constructed the co-expression regulatory network to detect the interaction between the TaZIP genes and miRNAs using a dataset of 173 RNA-seq, based on the weighted correlation of their expression.¹¹ *tae-miR164* and *tae-miR5384-3p* had major target genes, 10 and 6, respectively, while the *TaZIP25* and *TaZIP27* genes were targeted by the greatest number of *tae-miRNAs* (five *tae-miRNAs*; **Figure 3** and **Supplementary Table S4**). Other genes targeted by a major number of *tae-miRNAs* were *TaZIP20* (targeted by four *tae-miRNAs*), *TaIRT2-D*, *TaZIP13-A*, and *TaZIP14-D* (targeted by three *tae-miRNAs*). Three genes of *TaIRT2-A*, *TaZIP14-B*, and *TaZIP13-B* were targeted by two *tae-miRNAs*. **Figure 3**

¹¹<http://www.mirbase.org/>

demonstrates that miRNA164 targeted *TaIRT2-A*, miRNA9660-5p targeted *TaZIP13-A*, and tae-miR5084 targeted *TaZIP14-B*.

Expression Patterns of TaZIP Genes in Four Tissue and Grain Filling Stage

Overexpression of some ZIP genes may enhance zinc and iron content, thus enhancing grain and fruit quality. RNA-Seq data were downloaded for analyzed the tissue expression pattern. To investigate the expression pattern of the TaZIP genes in wheat grain, we designed 39 primers for the fluorescence quantification of wheat ZIP genes (Supplementary Table S5). In total, 39 ZIP genes are expressed in wheat four tissues. Most were highly expressed in the roots but less in the grain (Figure 4A and Supplementary Table S6). Four genes (*TaZIP4-B*, *TaZIP4-A*, *TaZIP6-D*, and *TaZIP14-B*) were highly expressed in the stem and three genes (*TaZIP19*, *TaZIP6-A*, and *TaZIP2-A*) were highly expressed in the leaves. Except for *TaZIP16-A*

and *TaZIP16-B*, group ZIIII was highly expressed in four tissues, particularly in the grain (Figure 4). There were also differential expressions pattern between homoeologs genes: *TaIRT2-A* and *TaIRT2-B* displayed high expression in wheat grain, while *TaIRT1-D* displayed low expression in wheat grain (Figure 4A and Supplementary Table S6).

We utilized gene expression levels at seven DAF as a control to better understand the expression pattern of TaZIP genes during the grain filling stage. The results showed that TaZIPs have significantly different expression levels (Figure 4B and Supplementary Table S6). Thirty-one TaZIP genes displayed downregulated expression, with the lowest expression levels at 28 DAF. Eight TaZIP genes were highly expressed, but the expression trend was diverse, *TaZTP7-B* had highest expression in 28 DAF. At the grain filling stage, *TaIRT2-A*, *D* were also highly expressed. *TaZIP14-B* and *TaZIP14-A* were unique genes in that their expression was upregulated at the grain filling stage. The expression level of *TaZIP14-B* greatly increased,

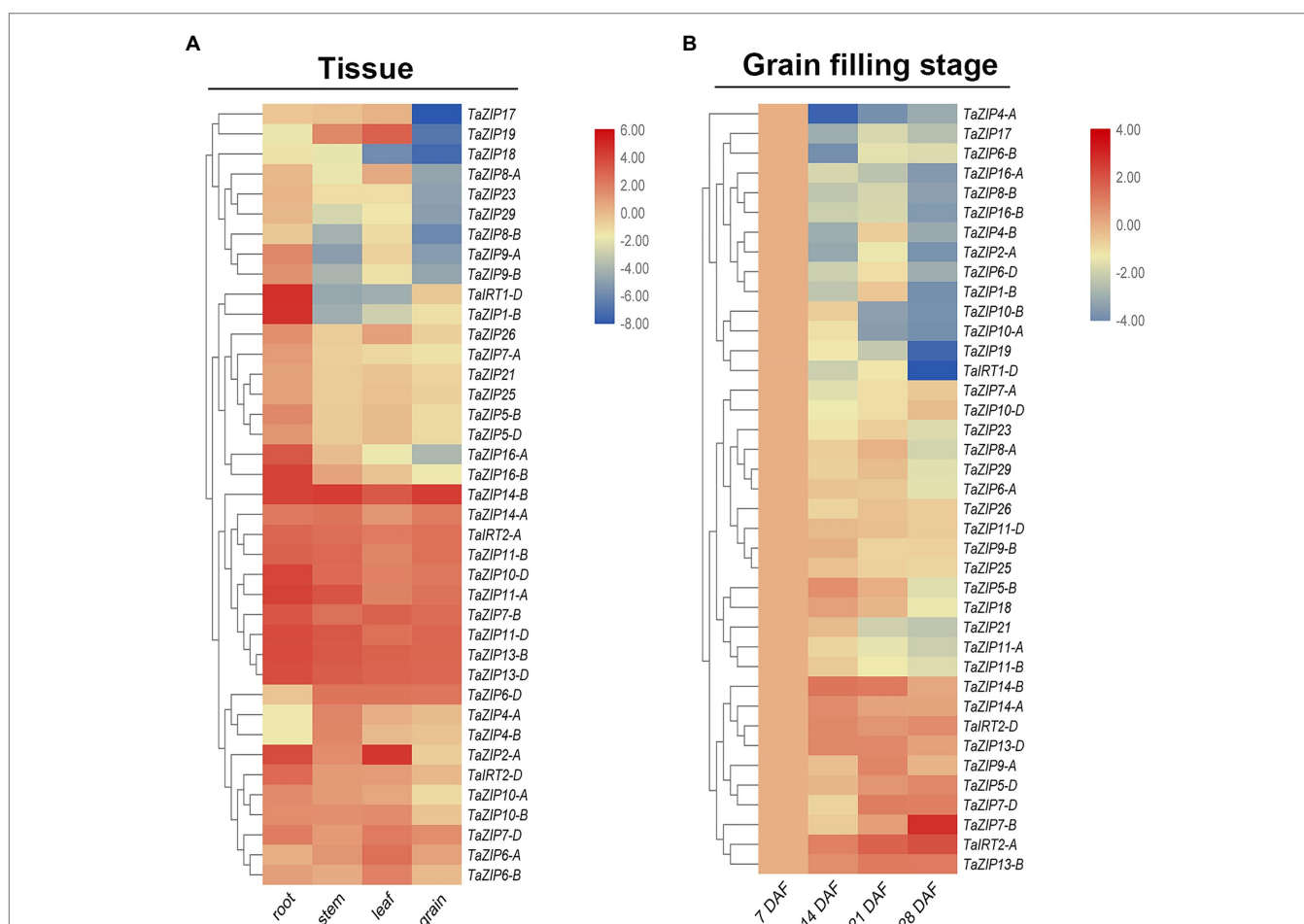


FIGURE 4 | Expression profiles of TaZIP genes in different tissues (A) and at different grain filling stages (B). (A) RNA-seq data of roots, leaves, stems, and grains of the genotype Chinese Spring were downloaded for the URGI database and used for expression profiles analysis. The expression level was determined by the fragments per kilobase per million (FPKM) calculated by StringTie v2.1.2 tool and log2 transformation was used for normalization. (B) The grain of genotype ZhongMai175 at different grain filling stages were collected and then used for qRT-PCR. The expression level was calculated according to the $2^{-\Delta\Delta Ct}$ method. Relative mRNA abundance of each gene was normalized with TaActin gene. DAF means day after flowering. The error bars indicate standard deviations. Three genes highlighted in yellow were used for cloned for downstream study.

while *TaZIP14-A* was expressed moderately. Both *TaZIP13-B* and *TaZIP13-D* were highly expressed, but their expression patterns were different. The expression levels of *TaZIP13-D* increased rapidly while *TaZIP13-B* displayed moderate expression (Figure 4B and Supplementary Table S6).

Expression Profiles of TaZIP Genes Under Zn and Fe Stress

We then analyzed the expression patterns of ZIP genes under different concentrations of ZnSO_4 and FeCl_3 solutions. Under low concentrations of ZnSO_4 stress, almost all ZIP genes were upregulated, but their expression trends differed. Sixteen ZIP genes were highly expressed (value more than 1) under 0.05 $\mu\text{mol/L}$ ZnSO_4 and decreased under 0.5 and 50 $\mu\text{mol/L}$ conditions (Figure 5A). The expression of these genes was suppressed when Zn concentration increased. Eighteen genes were highly upregulated under the concentration of 0.50 $\mu\text{mol/L}$ ZnSO_4 . Moreover, the expression of nine ZIP genes (*TaZIP21*,

-TaZIP5-D, *TaZIP5-B*, *TaZIP8-B*, *TaIRT1-D*, *TaIRT2-D*, *TaIRT2-A*, *TaZIP10-B*, and *TaZIP10-A*) increased in 50 $\mu\text{mol/L}$ ZnSO_4 . Especially *TaZIP14-B*, which displayed high levels of expression in the 0.5 $\mu\text{mol/L}$ solution, but low expressed in the 50 $\mu\text{mol/L}$ ZnSO_4 solution. The *TaZIP13-B* gene was moderately expressed in the 0.05 $\mu\text{mol/L}$ solution, while it was upregulated in other concentrations of ZnSO_4 solution (Figure 5A and Supplementary Table S6).

The expression pattern of TaZIPs in FeCl_3 was similar to ZnSO_4 . At low concentrations of FeCl_3 solution, the expression of all ZIP genes was increased, but of which 16 genes were further suppressed or moderately upregulated under 0.5 and 50 $\mu\text{mol/L}$ FeCl_3 treatment (Figure 5B and Supplementary Table S6). Furthermore, nine genes were highly expressed in the 0.05 and 0.5 $\mu\text{mol/L}$ FeCl_3 solutions. Eleven genes were highly expressed under 0.05 and 0.5 $\mu\text{mol/L}$ of FeCl_3 treatment. The genes *TaZIP14-B*, *TaZIP13-B*, and *TaZIP7-A*, *B*, *D* were all upregulated under high concentrations of FeCl_3 solution. The *TaIRT1-D* gene is key to transporting iron ions;

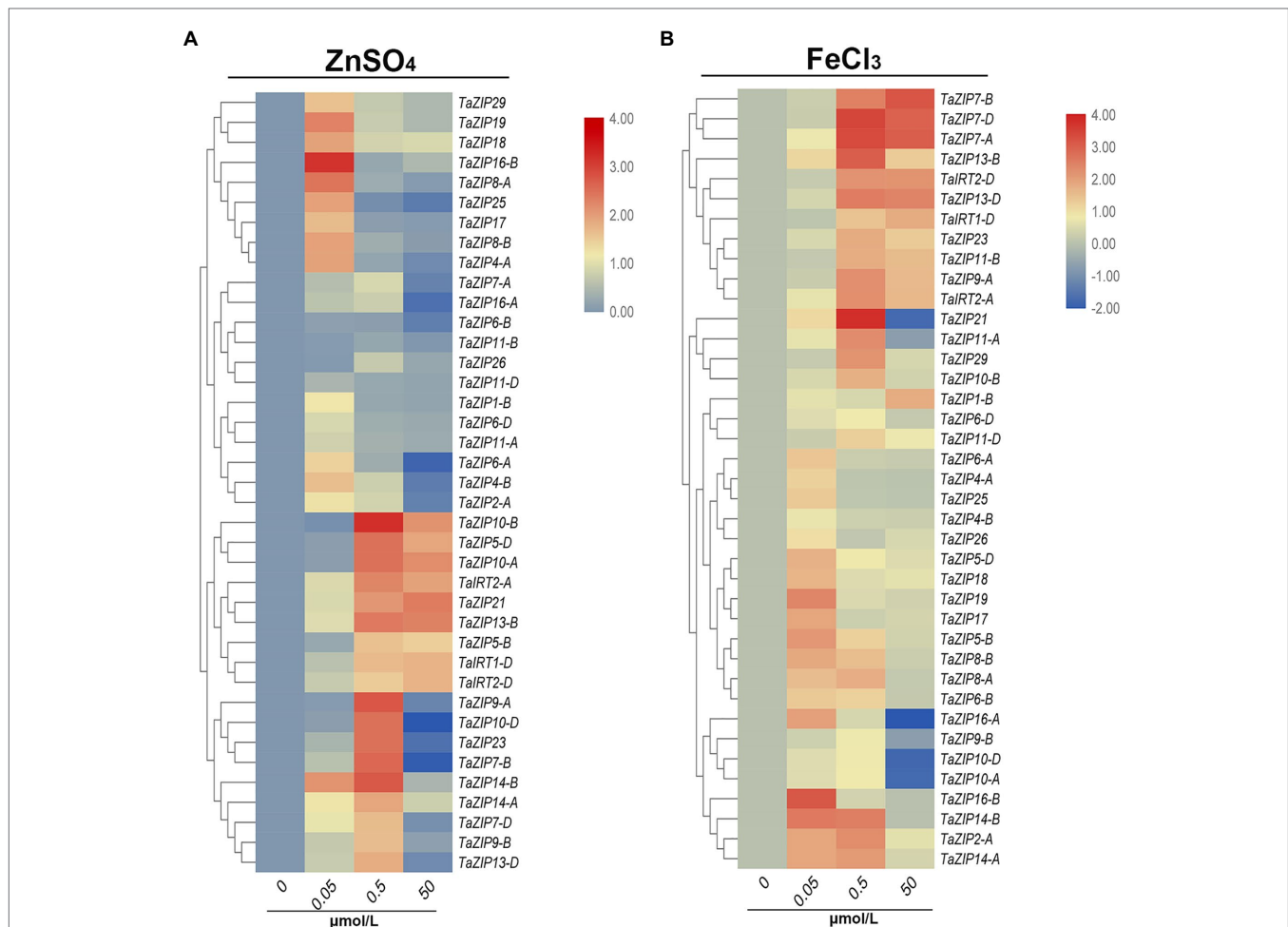


FIGURE 5 | Expression profiles of TaZIP genes under Zn or Fe stress through qRT-PCR analysis. The wheat seedling at two leaf stage of genotype Xiaobaimai were treated under standard nutrient condition (CK), 0.05, 0.5, 50 ZnSO_4 (A) and FeCl_3 (B) and the samples were harvested at 1 h after treatment. Data from qRT-PCR were analyzed according to the $2^{-\Delta\Delta C_t}$ method. Relative mRNA abundance of each gene was normalized with TaActin gene. The error bars indicate standard deviations.

however, its expression level under FeCl_3 treatment is lower than under ZnSO_4 treatment. Additionally, the expression pattern of *TaIRT2-A*, *D* was similar to that of *TaIRT1-D* (Figure 5B and Supplementary Table S6). We also performed Fe and Zn starvation treatment that wheat seedlings at two leaf stage were treated with Hoagland nutrient solution as CK and Hoagland medium lacking ZnSO_4 (Zn-deficient) or Fe (III)-EDTA (Fe-deficient). After treated by 6 and 12h, six genes were upregulated in roots and shoots and mainly expressed in roots (Supplementary Figure S3). *TaIRT2-A*, *D* were upregulated slightly under Zn-deficiency while highly expressed under Fe-deficiency. *TaZIP13-B*, *D* and *TaZIP14-A*, *B* were highly expressed under Zn-deficiency condition.

Expression Pattern of miRNA Under Zn and Fe Stress

To investigate whether miRNA degraded the ZIP genes, we analyzed the expression pattern of three miRNAs in wheat under ZnSO_4 and FeCl_3 solutions. These three miRNAs target *TaIRT2-A*, *TaZIP14-B*, and *TaZIP13-B* (Supplementary Tables S4 and S5).

Three miRNAs were slightly downregulated under 0.05 and $0.5\mu\text{mol/L}$ of the ZnSO_4 solution compared with the control (Figure 6). *tae-miR164* was also slightly downregulated in $50\mu\text{mol/L}$ ZnSO_4 solution, while *tae-miR5084* and *tae-miR395a* were slightly upregulated under $50\mu\text{mol/L}$ of the ZnSO_4 solution.

tae-miR164 displayed low levels of expression in the FeCl_3 solution compared with control, with the lowest expressed in $0.5\mu\text{mol/L}$ FeCl_3 solutions. *tae-miR395a* and *tae-miR5084* had similar expression patterns in the FeCl_3 solution. These two miRNAs were downregulated in the 0.05 and $0.5\mu\text{mol/L}$ FeCl_3 solutions, with the lowest expression in the $0.05\mu\text{mol/L}$ FeCl_3 solution. However, under $50\mu\text{mol/L}$ of FeCl_3 , these two miRNAs were upregulated slightly and did not differ from the control (Figure 6). The expression pattern of miRNAs contrary to the targeted genes.

Functional Analysis of Three TaZIPs by Complementation in Yeast Cells

After demonstrating that the three genes were upregulated when exposed to Zn and Fe stress, we also revealed the biological function of three ZIP genes (*TaZIP14-B*, *TaZIP13-B*, and *TaIRT2-A*) by yeast complementation analysis (Supplementary Table S5). *OsZIP3*, *OsZIP5*, and *OsIRT1* were chosen as positive controls, all of which have been demonstrated to be involved in Zn and Fe transport in rice (Chen et al., 2008; Tiong et al., 2014).

Three yeast strains wild-type DY1455, the *Saccharomyces eviscerate* *zrt1zrt2* mutant (ZHY3), and the *fet3fet4* mutant (DEY1453) were used, to verify that the three wheat TaZIP genes were capable of restoring the ability to transport zinc and iron in the mutant yeast. The full-length cDNA of both the wheat and rice genes were inserted and expressed in the two mutants. The transformed ZHY3 with TaZIP genes were grown on an SD medium with 0.4mM EDTA and the transformed

DEY 1453 were grown on a SD medium with 50mM MES. The results demonstrated that the growth of the ZHY3 yeast with pDR195 was inhibited under zinc-deficient conditions in a normal SD-Ura medium, while the mutant with TaZIP genes and rice genes successfully recovered from the growth defect (Figure 7A). The *TaZIP13-B* gene reversed the growth defect. When the ZHY3 yeast was exposed to a $200\mu\text{M}$ ZnSO_4 medium, the growth of ZHY3 was not inhibited. The growth of DEY1453 was similar to ZHY3 (Figure 7B). Under Fe-deficient conditions, the growth of DEY1453 containing a vector was severely inhibited, while the growth was reversed during the expression of TaZIP and rice genes. Once a sufficient amount of FeCl_3 was supplied, growth recovered. *TaZIP13-B* demonstrated the strongest propagation under Fe-deficient conditions. These results revealed that *TaZIP14-B*, *TaZIP13-B*, and *TaIRT2-A* could effectively complement the zinc transporter mutant *zrt1zrt2* and the iron transporter mutant *fet3fet4*, suggesting they could successfully transport Zn and Fe.

Gene Functional Analyzed by Arabidopsis thaliana

The *TaZIP13-B* gene was upregulated both at the grain filling stage and under Fe/Zn stress. The yeast complementation experiment proves that *TaZIP13-B* can transport Fe and Zn in a yeast mutant. Therefore, *TaZIP13-B* was transformed into *Arabidopsis* to verify this gene function. We selected lines OE-1, OE-2, and OE-3 with high expression levels for further analyses (Figure 8A).

On the MS, MS-Zn, and MS-Fe medium, the germination rate of the transgenic lines OE1, OE2, OE3, and WT approached 100%, with no significant difference between transgenic lines and WT (Figures 8B,C). When treated with $50\mu\text{mol/L}$ of Zn and Fe, the germination rate of four lines decreased, and no significant difference was observed between them (Figures 8B,C). When exposed to $200\mu\text{mol/L}$ of Zn and Fe MS medium, the germinations of the OE1-3 line were significantly higher than the WT line. The germination rate of OE1 was 85.3%, while that of OE2 was 82.7%, OE3 was 93.3%, and WT was 46.7% in $200\mu\text{mol/L}$ Fe MS medium. In $200\mu\text{mol/L}$ Zn MS, the germination rates of OE1-3 were 89.3, 90.7, and 85.3%, respectively, while that of WT was 74.7%. Compared with OE1 and OE3, the transgenic line OE2 had the highest germination on $300\mu\text{mol/L}$ Fe MS (up to 58.7%), while WT had a germination rate of only 22.7%. All three transgenic lines had significantly higher germination rates than WT line (Figures 8B,C).

Root length is another index used to evaluate plant tolerance to Zn and Fe stresses. We found no significant difference in root length between the three transgenic lines and WT on MS, Zn-, Fe-, and $50\mu\text{mol/L}$ Zn/Fe MS medium (Figures 8D,E); however, the number of roots in Zn-, Fe-, and $50\mu\text{mol/L}$ Zn/Fe MS medium increased. When transgenic lines and the WT line were placed on $200\mu\text{mol/L}$ Fe MS medium, the root lengths were significantly inhibited, but the root lengths of the three transgenic lines were significantly longer than WT line. The root length of OE2 was 1.78 cm, which was the

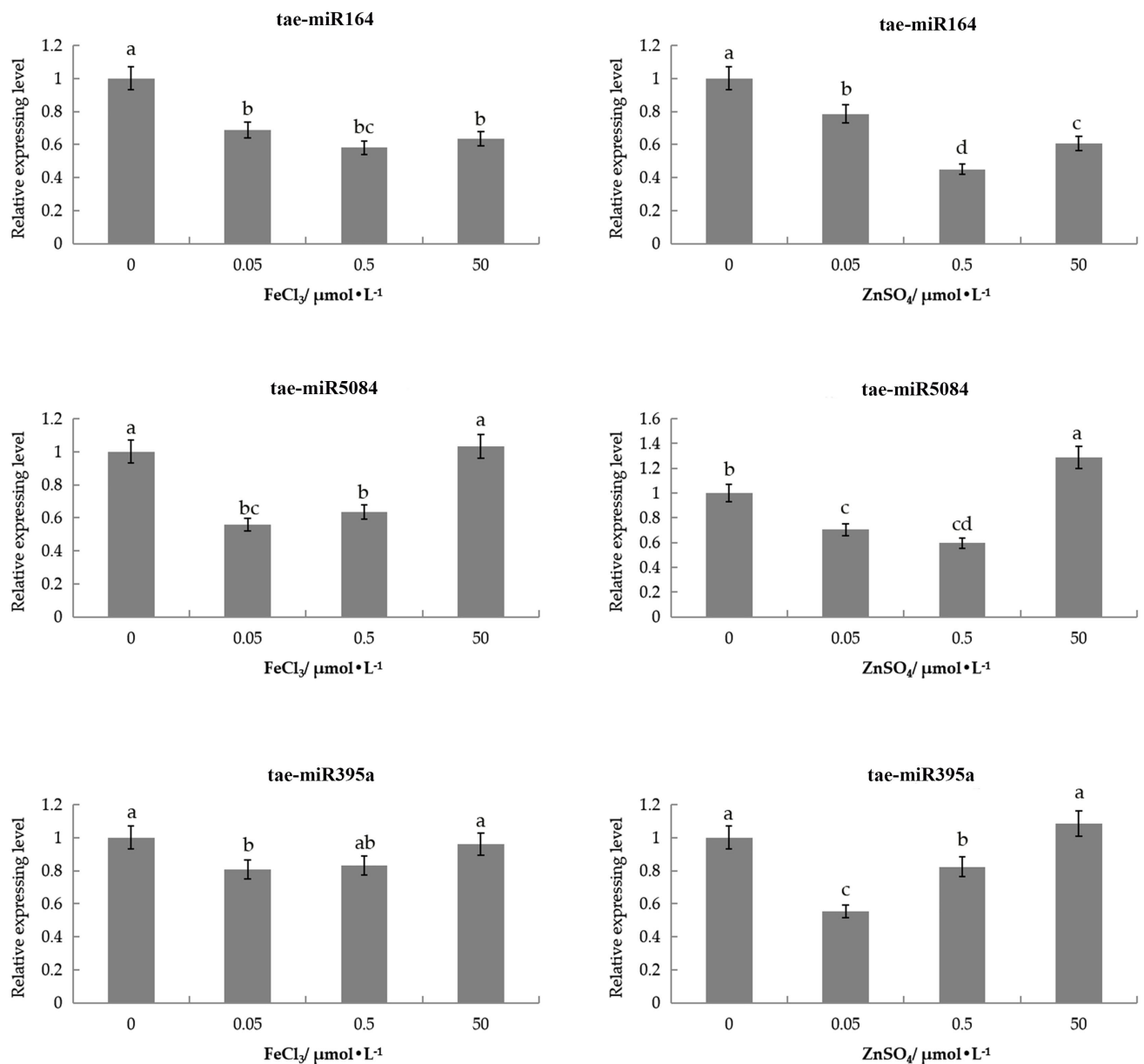


FIGURE 6 | Expression patterns of three taе-miRNAs under Zn and Fe stress through qRT-PCR analysis. The expression level was calculated according to the $2^{-\Delta\Delta Ct}$ method. Relative miRNA abundance of each gene was normalized with the expression of 0h. Error bars indicate the standard deviations, and different letters are significantly different.

longest root length on the 200 μmol/L Fe MS medium. On the 200 μmol/L Zn MS medium, the three transgenic lines were significantly longer than the WT line. On the 300 μmol/L Fe MS, the root lengths were further inhibited compared to the 200 Fe MS medium, and the root lengths of the three transgenic lines were significantly longer than the WT line. The root lengths were also inhibited in 300 MS Zn MS, though the root lengths of the three transgenic lines were significantly longer than that of the WT line (**Figures 8D,E**).

To further evaluate the tolerance of the three transgenic lines and WT to Zn and Fe stress, we treated the transgenic

and WT lines with different concentrations of a ZnSO₄ and FeCl₃ solution. Under 200 μmol/L FeCl₃, the WT leaves wilted and yellowish-brown spots appeared on a few leaves, though this did not occur on the OE1, OE2, and OE3 lines (**Figure 8F**). Under 300 and 400 μmol/L of the FeCl₃ solution, most WT leaves were brown while transgenic lines were normal. Almost all of the WT petiole browned, particularly under 400 μmol/L FeCl₃ solution treatment (**Figure 8E**). We then measured the chlorophyll content of four lines. As the concentration of the FeCl₃ solution increased, the chlorophyll content of the four lines decreased (**Figures 8G,H**). However, the chlorophyll content

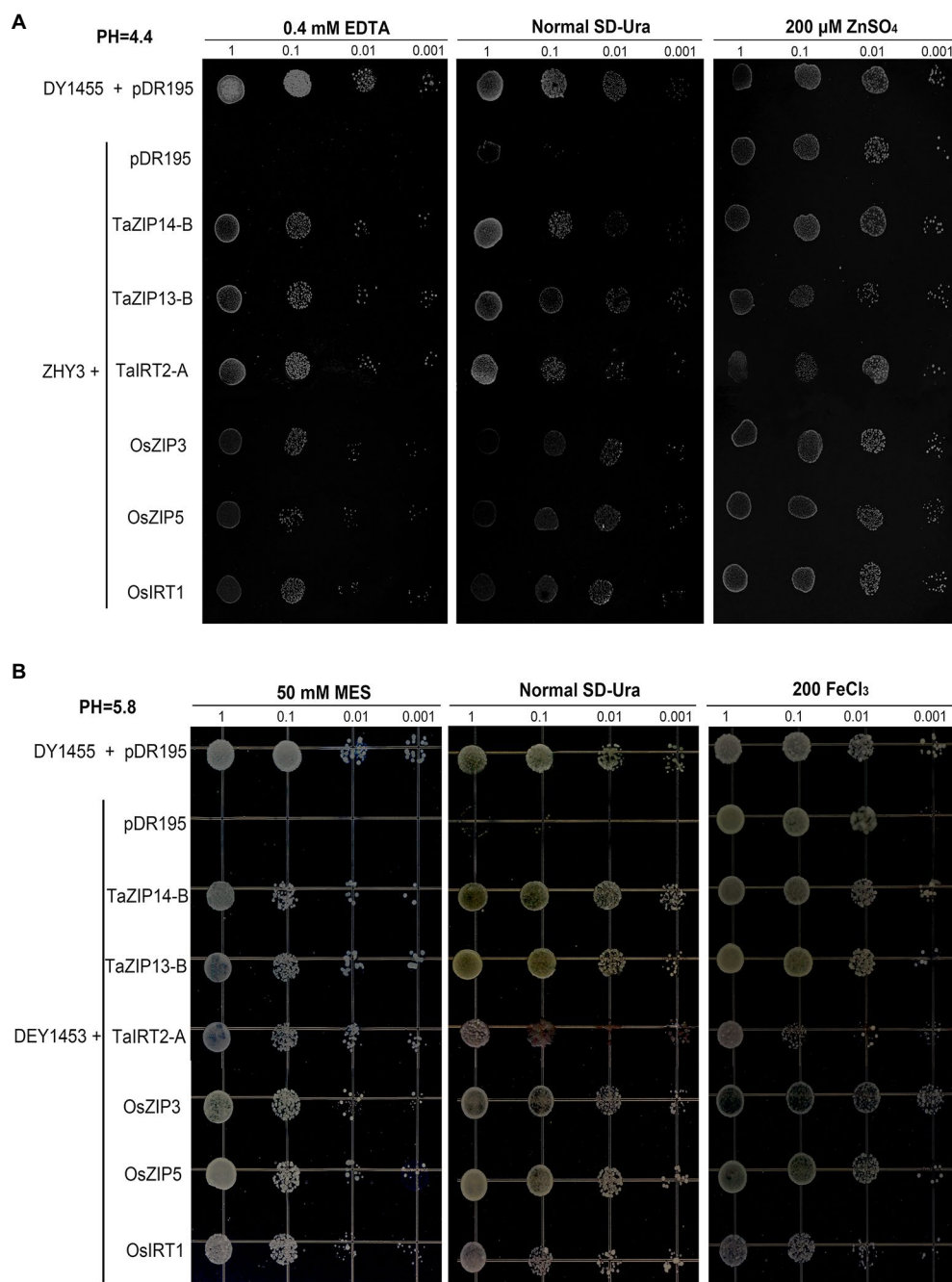
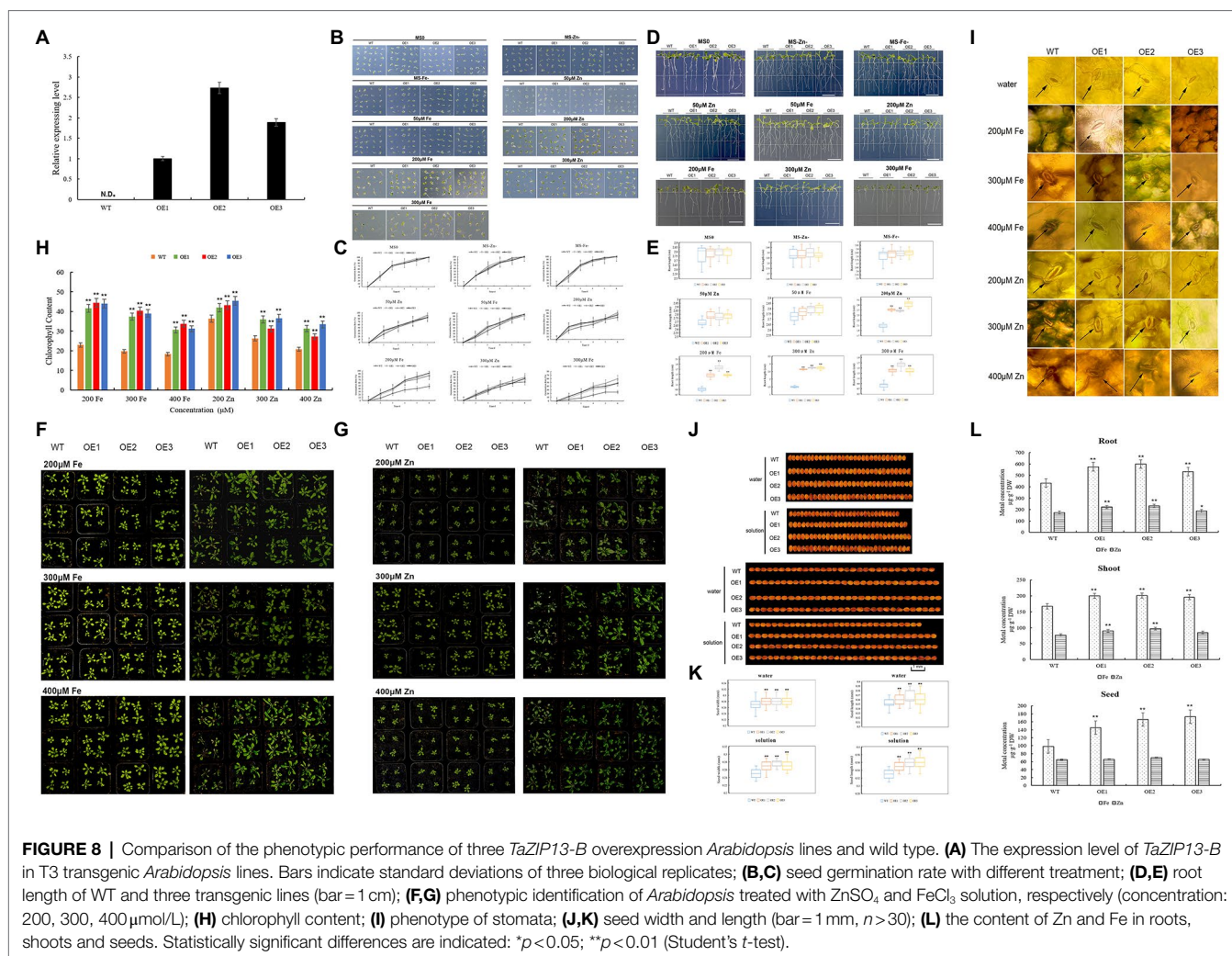


FIGURE 7 | Functional complementation of yeast Zn and Fe transport mutants by TaZIPs under different pH conditions. **(A)** The Zn transport mutant *zrt1zrt2* (pH 4.4); **(B)** the Fe transport mutant *fet3fet4* (pH 5.5–5.8). Mutant transformed with the expression vector pDR195 carrying *TaIRT2-A*, *TaZIP14-B*, or *TaZIP13-B* or a functionally characterized ZIP gene, *OsZIP5*, *OsZIP8*, or *OsIRT1*. The wild-type (WT) strain DY1455 transformed with pDR195 was used as a positive control, and the yeast *zrt1zrt2* or *fet3fet4* mutant transformed with the empty vector pDR195 was used as a negative control. The transformed yeast cells were grown under different metal conditions as indicated, and the transformed *fet3fet4* was grown on medium with pH 5.8. Cell concentration was adjusted to $OD_{600} = 1$ and serial dilutions (1.0, 0.1, 0.01, and 0.001) were made. For assay, 5 μ l of each dilution was spotted on plates and grown for 6 days at 30°C.

in the three transgenic lines was significantly higher than in the WT line. The three transgenic lines exhibit greater $FeCl_3$ resistance than the WT line, according to observations of the stomata in all four lines (**Figure 8I**). The stomata opened on all four lines when treated with water. Under 200 $FeCl_3$ μ mol/L

treatment, the stomata of the WT line closed slightly, while the stomata of the three overexpression lines opened. When treating the four lines with 300 μ mol/L $FeCl_3$ solutions, the stomata of the WT closed while OE1, OE2, and OE3 remained open (**Figure 8I**). Under 400 μ mol/L $FeCl_3$ solution, the stomata



of all lines closed, and the stomata of the WT died. When treated with 200 $\mu\text{mol/L}$ ZnSO_4 solution, the leaves of the WT line turned gray, while the leaves of the three transgenic lines remained green (**Figure 8G**). The results were similar to the 200 $\mu\text{mol/L}$ ZnSO_4 treatment when they were treated with the 300 and 400 $\mu\text{mol/L}$ ZnSO_4 solution. However, when treated with the 300 and 400 $\mu\text{mol/L}$ ZnSO_4 solutions, part leaves in the transgenic lines turned gray and wilted (**Figure 8G**). The chlorophyll content of both of the transgenic and WT plants was analyzed. Under treatment with three different concentrations of ZnSO_4 solution, the chlorophyll content in the WT line was significantly lower than in the transgenic lines (**Figure 8H**). The stomata phenotype of all lines under the ZnSO_4 treatment was similar to that of the FeCl_3 treatment (**Figure 8I**).

Zn is important to photosynthesis in plants, and photosynthesis is related to crop yield. Therefore, we analyzed whether the *TaZIP* genes affect seed size. When the four lines were treated with water, the seed width and length of the three transgenic lines were longer than that of the WT line (**Figures 8J,K**). When treated with 200 $\mu\text{mol/L}$ FeCl_3 and ZnSO_4 solution, all lines' seed size shrank, but the three transgenic lines' seed

breadth and length remained somewhat longer than the WT line (**Figures 8J,K**). Our results indicate that transferring the wheat *TaZIP13-B* gene into *Arabidopsis* increases seed size and might increase production.

We used 0.15g samples of the roots, shoots, and seeds to measure the metal contents in this study. Compared with the WT, the three overexpression lines accumulated more Fe (23.6–38% higher) and Zn (7.5–33% higher) in the roots, while overexpression lines also accumulated more Fe (17.3–20.3%) and Zn (10.3–26.0%) in the shoots than the WT line. Compared with the WT, the seeds of the transgenic lines have a higher Fe content and a higher Zn content (**Figure 8L**). These results indicate that transgenic lines can absorb more Fe and Zn from the soil, enriching the Fe and Zn in seeds.

DISCUSSION

Zinc and iron are two microelements that are essential for plant development. Inadequate zinc and iron can result in etiolation, wilting, and even death (Vert et al., 2001). The

primary reason for zinc deficiency in plants is soil with low levels of Zn and Fe (Almendros et al., 2013). Approximately 30% of the world's agricultural area is Zn-deficient, which affects both grain yield and the Zn concentration in grains (Paul, 2015). To achieve sustained Zn uptake from the environment, plants have a dual-transporter system that includes high- and low-affinity Zn transporters called ZIPs (transporter-like protein; Eide, 2006). This protein family has been reported in many species, including *Arabidopsis*, rice, barley, maize, and wheat (Bughio et al., 2002; Pedas et al., 2008; Li et al., 2013; Tiong et al., 2015; Evens et al., 2017). Previous studies have identified 42 ZIP genes in wheat, while many ZIP genes have not yet been identified (Tiong et al., 2015; Evens et al., 2017). In this study, we identified 58 ZIP genes in wheat, which includes 42 previously identified ZIP genes. Additionally, we analyzed the expression pattern both of specific tissues and under $\text{ZnSO}_4/\text{FeCl}_3$ treatment. We also analyzed the ZIP gene structure and the motifs of the TaZIPs. These genes were distributed on all chromosomes, except for chromosome 5. The localization of the ZIP genes was uneven, which could be due to the specific retention and dispersal of TaZIPs during polyploidization. The sequence length of wheat ZIP genes varied significantly, while the transmembrane domain between III and IV can be changed (Guerinot, 2000). The subcellular localization of the most TaZIPs proteins was predicted to be located on the membrane. Our results were consistent with those of ZmZIPs, AtZIPs, and HvZIPs (Lin et al., 2009; Li et al., 2013; Tiong et al., 2014). The plasma membrane is an important region for Zn and Fe transport since plant proteins located on the plasma membrane can quickly assimilate Zn and Fe from the environment (Schneider, 1983). Other ZIP proteins are located on the vacuolar membrane, including *AtZIP1* and *OsZIP6*. miRNA is a regulatory factor that plays an important role in regulating the expression level of plant proteins after transcription (Vaucheret, 2006). Therefore, we constructed the network of miRNAs and target genes and found that tae-miR164 and tae-miR5084 had the most targeted genes. These two miRNAs were targeted to TaIRT genes, in particular, tae-miR164 was targeted to *TaIRT1-A* and *TaIRT2-A, B, D*. This suggests that tae-miR164 and tae-miR5084 could each play an important role in the uptake and enrichment of Fe from the environment.

Most TaZIP genes are primarily expressed in the roots, while others are expressed in the leaves or stems. Our results demonstrate that the expression of most TaZIP genes in the roots helps absorb and transport Zn and Fe. Several studies have revealed that Zn and Fe were primarily absorbed by the roots and delivered to different tissues through the phloem-tropic mode (Yamaji and Ma, 2014). The Zn and Fe contents in grain is one of the most important indexes measuring wheat quality (Ziaeeian and Malakouti, 2001). Zn and Fe accumulation typically occurs in the grain during the grain-filling stage (Tavarez et al., 2015). Some studies have demonstrated that mineral deficiency can induce the overexpression of ZIP genes (Li et al., 2013). Other studies have reported the relationship between the Zn and Fe content the overexpression of ZIP genes in cereal (Lee and An, 2009; Tiong et al., 2014). This

study found that nine genes were upregulated at the grain-filling stage, indicating that these genes are likely involved in Zn and Fe accumulation in grain. *TaIRT2* expression levels were upregulated and *TaIRT1* expression levels were downregulated, which is similar to *OsIRT1* and *OsIRT2* (Nakanishi et al., 2010). In most plants, *IRT1* and *IRT2* have different transport substrates and different expression patterns during the plant growth stage (Vert et al., 2001; Pedas et al., 2008).

Most ZIP genes are upregulated under Zn- and Fe- deficient conditions (Mäser et al., 2001; Mizuno et al., 2008; Tiong et al., 2015; Evens et al., 2017). In *Arabidopsis*, *AtZIP1-5*, *AtZIP9-12*, and *AtIRT3* were induced by Zn-deficiency treatment; in rice, *OsIRT1* and *OsIRT2* were induced by Fe-deficiency treatment; and in wheat, *TaZIP3*, -5, -7, and -13 were induced by Zn-deficiency treatment (Evens et al., 2017). The ZIP transporter is a dual-transporter system, which includes high-affinity and low-affinity Zn transporters (Sillanpää and AGL, 1982; Mizuno et al., 2008). The high-affinity system is saturated at approximately $0.1 \mu\text{mol/L}$, while the low-affinity system shows a linear relationship that varies from concentrations of 0.5 to $50 \mu\text{mol/L}$ (Lee et al., 2010a,b). The Zn uptake system is a dual system in wheat (Reid et al., 1996). Our study demonstrated that approximately half of the TaZIP genes were highly expressed in the $0.05 \mu\text{mol/L}$ ZnSO_4 treatment, while others were highly expressed in the $0.5 \mu\text{mol/L}$ ZnSO_4 treatment. The TaZIP genes displayed a similar expression pattern in FeCl_3 solution. Nineteen TaZIPs were highly expressed under $0.05 \mu\text{mol/L}$ FeCl_3 and 23 TaZIPs were highly expressed in $0.5 \mu\text{mol/L}$ FeCl_3 . A previous study found that expression patterns of ZIP genes differed under different concentrations of Zn and Fe treatment. Based on Zn affinity, we considered the 16 TaZIP genes with the highest expression in the $0.05 \mu\text{mol/L}$ ZnSO_4 treatment to be high-affinity Zn transporters, while other TaZIP genes were considered low-affinity Zn transporters. In this study, we also found that four TaZIPs (*TaZIP18*, *TaZIP4-B*, *TaZIP29*, and *TaIRT2-A*) were upregulated under the 0.05 and $0.5 \mu\text{mol/L}$ ZnSO_4 treatments. We also prove that six ZIP genes were upregulated under Zn and Fe deficient conditions.

The regulatory role of miRNA inhibiting the expression of target genes, meaning that miRNAs and target genes have opposing expression patterns (Zamore et al., 2000; Bernstein et al., 2001). However, recent research has showed that miRNA also can activate gene transcription (Xiao et al., 2017). In this study, we analyzed the expression of three miRNAs under Zn and Fe stress. Our results demonstrated that three miRNAs could downregulated under low concentrations of ZnSO_4 and FeCl_3 . However, tae-miR5084 and tae-miR395a were upregulated in $50 \mu\text{mol/L}$ of ZnSO_4 and FeCl_3 . Absorbing excessive Fe and Zn is toxic to plants, meaning that wheat may upregulate miRNA to inhibit TaZIP gene expression. The overexpression of tae-miR399-A1 could inhibit the expression of the *TaPHO2-A1*, *B1*, *D1* genes in a high-phosphorus aqueous solution, but wheat accumulates more Pi in its leaves (Ouyang et al., 2016). In this study, tae-miR5084 and tae-miR395a were both upregulated in $50 \mu\text{mol/L}$ ZnSO_4 and FeCl_3 , inhibiting the expression of targeted genes.

In yeast, the high-affinity transporter gene (*Zrt1*) is responsible for the uptake of Zn in a Zn-deficient medium. When Zn is abundant, *Zrt1* is repressed and the low-affinity transporter (*Zrt2*) mediates Zn uptake (Eide, 2006). ZHY3 is a yeast mutant that lack the *zrt1* and *zrt2* genes and unable grow on SD media without ZnSO₄. *fet3fet4* DEY1453 is another mutant that cannot normally grow on the SD media without FeCl₃. This growth deficiency may be reversed by inserting a functioning gene into these mutants. Yeast complementation has been used to demonstrate that ZIP genes can reverse growth defects in the *zrt1zrt2* and *fet3fet4* double mutant (Mäser et al., 2001). In this study, *TaZIP14-B*, *TaIRT2-A*, and *TaZIP13-B* inserted into the yeast mutant, and yeast complementation assays demonstrated that these genes could reverse the growth defect. Our results demonstrated that these three genes could effectively transport Zn and Fe. While some wheat ZIP genes have been studied, none of these three genes have been tested (Evens et al., 2017; **Supplementary Figure S1**).

Plants have evolved two methods of avoiding toxic metals. The first is to exclude metal from the plant, and the second is to enrich the metal elements in a particular organelle. In plants, the roots are responsible for the uptake of metal elements, while the vacuoles are responsible for their exclusion and enrichment. This process involves YSL genes, CDF genes, and ZIP genes (Colangelo and Guerinot, 2006). Aside from transporting zinc and iron, ZIP transporters also transport other metals, including Cd, Ni, and Mn. Most ZIP transporters enhance Zn and Fe at the root when they are expressed, however ZIP gene expression may also enrich Zn and Fe in the stem and leaves (Salt et al., 1995). Fe and Zn are dynamically balanced in plants. Exposing a plant to high concentrations of metal elements destroys the balance between the production and scavenging of free radicals in its cells, which produces a large number of reactive oxygen radicals and induces the peroxidation of unsaturated fatty acids in the membrane. It also causes heavy metal poisoning in plants (Bernstein et al., 2001; Breusegem and Dat, 2006; Zhao, 2007). In this study, we treated three *TaZIP13-B* transgenic lines and one WT line with different concentrations of FeCl₃ solutions and found that the leaves of the WT developed brown spots and most petioles died. Under ZnSO₄ treatment, the WT leaves turned gray and the three overexpression lines remained normal. Heavy metals are primarily toxic to plants because they inhibit chlorophyll synthesis, affecting photosynthesis and inducing chlorosis of the leaves (Breusegem and Dat, 2006). *OsIRT1* overexpression also results in less chlorosis in transgenic plants under Fe-deficient conditions (Lee and An, 2009). In this study, the chlorophyll content in the WT line was significantly lower than in the transgenic lines when exposed to Fe and Zn solutions.

The hormone indole-3-acetic acid (IAA) is related to lateral root formation. Previous studies found that IAA levels increased under Cu and Cd stress but there was no significant change in the roots compared with the control. Zn stress caused significant increases in root branching (Sofa et al., 2013). In this study, the root branch increased under 50 μmol/L Zn/Fe stress due to increases in the IAA concentration, resulting in

lateral formation. Under MS Zn-/Fe- conditions, the number of lateral roots also increased. Previous studies found that the lateral root of *Arabidopsis* increased under Pi- and Fe-deficient conditions (Rai et al., 2015), while there is no evidence to prove that Zn- or Fe-deficiency promotes the development of lateral roots in *Arabidopsis*. *Arabidopsis* generated different patterns of root system architecture when subjected to different combinations of Pi, nitrate (N), potassium (K), and sulfate (S) deficiencies (Kellermeier et al., 2014). This indicates that *Arabidopsis* has an innate ability to integrate and translate multiple nutrient deficiencies into a complex root developmental program.

Zn and Fe are vital for plant growth and are related to dry matter accumulation in plants, when plants reach the reproductive stage, their photosynthetic products accumulate in the grain. Therefore, the size of the seed is related to the accumulation of dry matter in the early stage of plants (Cock and Yoshiida, 1972; Yoshida, 1972; Hirose et al., 2008). However, this increase in seed size is due to overexpression in the plant body, not by seed-restricted expression. This indicates that seed enlargement is due to overexpression in vegetative organs such as the leaves (Hakata et al., 2012). In this study, the seeds of three overexpression lines of *Arabidopsis* were larger than the WT line, indicating that the chlorophyll content in transgenic lines is higher than in WT lines.

We also detected the Fe and Zn content in plant tissues. Previous studies found that the Zn and Fe content in seeds improved when *ZmZIP7*, *ZmZIP3* and *ZmZIP5* and *ZmIRT1* were transferred to wild *Arabidopsis* and maize (Li et al., 2013, 2016, 2019b). Our results demonstrated that the Fe and Zn contents in the roots and shoots were more enriched in overexpression lines than in the WT line and that Fe content was particularly increased in the seeds. This study indicated that *TaZIP13-B* can enrich and transport Fe and Zn in transgenic lines and improve Fe and Zn content in seeds.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

WZ and SL conceived and designed the study. SL, ZL, and HL collected the samples. LG, HL, XN, and SC analyzed the data. SL wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Transgenic Key project of Ministry of Agriculture of China (2020ZX08009-15B), Key project of Research and Development Program of Shaanxi

(2019NY-014), and the National Natural Science Foundation of China (grant no. 31871611).

ACKNOWLEDGMENTS

We thank Researcher Rumei Chen (Agricultural Biotechnology Institute, China) for providing the yeast strains DEY1453, ZHY3, and DY1455.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.748146/full#supplementary-material>

REFERENCES

- Ajeesh Krishna, T. P., Maharajan, T., Victor Roch, G., Ignacimuthu, S., and Antony Ceasar, S. (2020). Structure, function, regulation and phylogenetic relationship of ZIP family transporters of plants. *Front. Plant Sci.* 11:662. doi: 10.3389/fpls.2020.00662
- Almendros, P., Gonzalez, D., and Alvarez, J. M. (2013). Residual effects of organic Zn fertilizers applied before the previous crop on Zn availability and Zn uptake by flax (*Linum usitatissimum*). *J. Plant Nutr. Soil Sci.* 176, 603–615. doi: 10.1002/jpln.201100333
- Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363–366. doi: 10.1038/35053110
- Breusegem, F. V., and Dat, J. F. (2006). Reactive oxygen species in plant cell death. *Plant Physiol.* 141, 384–390. doi: 10.1104/pp.106.078295
- Briat, J. F., and Lebrun, M. (1999). Plant responses to metal toxicity. *C. R. Acad. Sci. III* 322, 43–54. doi: 10.1016/s0764-4469(99)80016-x
- Bughio, N., Yamaguchi, H., Nishizawa, N. K., Nakanishi, H., and Mori, S. (2002). Cloning an iron-regulated metal transporter from rice. *J. Exp. Bot.* 53, 1677–1682. doi: 10.1093/jxb/erf004
- Chen, W. R., Feng, Y., and Chao, Y. E. (2008). Genomic analysis and expression pattern of OsZIP1, OsZIP3, and OsZIP4 in two rice (*Oryza sativa* L.) genotypes with different zinc efficiency. *Russ. J. Plant Physiol.* 55, 400–409. doi: 10.1134/S1021443708030175
- Cock, J. H., and Yoshiida, S. (1972). Accumulation of ¹⁴C-labelled carbohydrate before flowering and its subsequent redistribution and respiration in the rice plant. *Jpn. J. Crop. Sci.* 41, 598–607. doi: 10.1626/jcs.41.226
- Colangelo, E. P., and Gueriot, M. L. (2006). Put the metal to the petal: metal uptake and transport throughout plants. *Curr. Opin. Plant Biol.* 9, 322–330. doi: 10.1016/j.pbi.2006.03.015
- Dai, X., and Zhao, P. X. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 39, W155–W159. doi: 10.1093/nar/gkr319
- Eide, D. J. (2006). Zinc transporters and the cellular trafficking of zinc. *Biochim. Biophys. Acta* 1763, 711–722. doi: 10.1016/j.bbamcr.2006.03.005
- Eide, D., Broderius, M., Fett, J., and Gueriot, M. L. (1996). A novel iron-regulated metal transporter from plants identified by functional expression in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 93, 5624–5628. doi: 10.1073/pnas.93.11.5624
- Evans, N. P., Buchner, P., Williams, L. E., and Hawkesford, M. J. (2017). The role of ZIP transporters and group F bZIP transcription factors in the Zn-deficiency response of wheat (*Triticum aestivum*). *Plant J.* 92, 291–304. doi: 10.1111/tpj.13655
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344
- Gietz, R. D., and Schiestl, R. H. (1995). Transforming yeast with DNA. *Mol. Cell. Biol.* 5, 255–269.
- Grotz, N., and Gueriot, M. L. (2006). Molecular aspects of Cu, Fe and Zn homeostasis in plants. *Biochim. Biophys. Acta* 1763, 595–608. doi: 10.1016/j.bbamcr.2006.05.014
- Supplementary Figure S1** | Phylogenetic relationship and homoeologs of the 58 TaZIP proteins. The gene that ends with S and L was identified by previous study. Genes highlighted in green were used for yeast complementation. Genes highlighted in yellow were used in this study by yeast complementation.
- Supplementary Figure S2** | Chromosome locations of these 58 TaZIP genes.
- Supplementary Figure S3** | Expression patterns of three ZIP genes under Zn- and Fe-deficient conditions. Seedling two leaf of Xiaobaimai shoots (S) and roots (R), under standard nutrient condition (CK), Zn-, Fe-deficiency treated, were harvested, respectively, at 0, 6, and 12 h, after treatment. (A) Treat with Zn-deficiency; (B) treat with Fe-deficiency. Data from real-time RT-PCR experiments were analyzed according to the 2^{-ΔΔC_t} method. Relative mRNA abundance of each gene was normalized with TaActin gene. The error bars indicate standard deviations.
- Gueriot, M. L. (2000). The ZIP family of metal transporters. *Biochim. Biophys. Acta* 1465, 190–198. doi: 10.1016/s0005-2736(00)00138-3
- Hakata, M., Kuroda, M., Ohsumi, A., Hirose, T., and Yamakawa, H. (2012). Overexpression of a rice TIFY gene increases grain size through enhanced accumulation of carbohydrates in the stem. *Biosci. Biotechnol. Biochem.* 76, 2129–2134. doi: 10.1271/bbb.120545
- Hansen, T. H., de Bang, T. C., Laursen, K. H., Pedas, P., Husted, S., and Schjoerring, J. K. (2013). Multielement plant tissue analysis using ICP spectrometry. *Methods Mol. Biol.* 953, 121–141. doi: 10.1007/978-1-62703-152-3_8
- He, Z. H., Chen, X. M., Wang, D. S., Zhang, Y., Xiao, Y. G., Li, F. J., et al. (2015). Characterization of wheat cultivar zhongmai 175 with high yielding potential, high water and fertilizer use efficiency, and broad adaptability. *Sci. Agric. Sin.* 48, 3394–3403. doi: 10.3864/j.issn.0578-1752.2015.17.007
- Hirose, T., Endler, A., and Ohsugi, R. (2008). Gene expression of enzymes for starch and sucrose metabolism and transport in leaf sheaths of rice (*Oryza sativa* L.) during the heading period in relation to the sink to source transition. *Plant Prod. Sci.* 2, 178–183. doi: 10.1626/pp.2.178
- Ishimaru, Y., Kim, S., Tsukamoto, T., Oki, H., Kobayashi, T., Watanabe, S., et al. (2007). Mutational reconstructed ferric chelate reductase confers enhanced tolerance in rice to iron deficiency in calcareous soil. *Proc. Natl. Acad. Sci. U. S. A.* 104, 7373–7378. doi: 10.1073/pnas.0610555104
- Ishimaru, Y., Suzuki, M., Tsukamoto, T., Suzuki, K., Nakazono, M., Kobayashi, T., et al. (2006). Rice plants take up iron as an Fe³⁺-phytosiderophore and as Fe²⁺. *Plant J.* 45, 335–346. doi: 10.1111/j.1365-3113X.2005.02624.x
- Itai, R. N., Ogo, Y., Kobayashi, T., Nakanishi, H., and Nishizawa, N. K. (2013). Rice genes involved in phytosiderophore biosynthesis are synchronously regulated during the early stages of iron deficiency in roots. *Rice* 6:16. doi: 10.1186/1939-8433-6-16
- Kambe, T., Yamaguchi-Iwai, Y., Sasaki, R., and Nagao, M. (2004). Overview of mammalian zinc transporters. *Cell. Mol. Life Sci.* 61, 49–68. doi: 10.1007/s00018-003-3148-y
- Kavitha, P. G., Kuruvilla, S., and Mathew, M. K. (2015). Functional characterization of a transition metal ion transporter, OsZIP6 from rice (*Oryza sativa* L.). *Plant Physiol. Biochem.* 97, 165–174. doi: 10.1016/j.plaphy.2015.10.005
- Kellermeier, F., Armengaud, P., Seditas, T. J., Danku, J., Salt, D. E., and Amtmann, A. (2014). Analysis of the root system architecture of *Arabidopsis* provides a quantitative readout of crosstalk between nutritional signals. *Plant Cell* 26, 1480–1496. doi: 10.1105/tpc.113.122101
- Krämer, U., Talke, I. N., and Hanikenne, M. (2007). Transition metal transport. *FEBS Lett.* 581, 2263–2272. doi: 10.1016/j.febslet.2007.04.010
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Lee, S., and An, G. (2009). Over-expression of OsIRT1 leads to increased iron and zinc accumulations in rice. *Plant Cell Environ.* 32, 408–416. doi: 10.1111/j.1365-3040.2009.01935.x
- Lee, S., Jeong, H. J., Kim, S. A., Lee, J., Gueriot, M. L., and An, G. (2010a). OsZIP5 is a plasma membrane zinc transporter in rice. *Plant Mol. Biol.* 73, 507–517. doi: 10.1007/s11103-010-9637-0

- Lee, S., Kim, S. A., Lee, J., Guerinot, M. L., and An, G. (2010b). Zinc deficiency-inducible OsZIP8 encodes a plasma membrane-localized zinc transporter in rice. *Mol. Cell* 29, 551–558. doi: 10.1007/s10059-010-0069-0
- Li, Q., Chen, L., and Yang, A. (2019a). The molecular mechanisms underlying iron deficiency responses in rice. *Int. J. Mol. Sci.* 21:43. doi: 10.3390/ijms21010043
- Li, S., Liu, X., Zhou, X., Li, Y., Yang, W., and Chen, R. (2019b). Improving zinc and iron accumulation in maize grains using the zinc and iron transporter ZmZIP5. *Plant Cell Physiol.* 60, 2077–2085. doi: 10.1093/pcp/pcz104
- Li, S., Zhou, X., Huang, Y., Zhu, L., Zhang, S., Zhao, Y., et al. (2013). Identification and characterization of the zinc-regulated transporters, iron-regulated transporter-like protein (ZIP) gene family in maize. *BMC Plant Biol.* 13:114. doi: 10.1186/1471-2229-13-114
- Li, S., Zhou, X., Zhao, Y., Li, H., Liu, Y., Zhu, L., et al. (2016). Constitutive expression of the ZmZIP7 in *Arabidopsis* alters metal homeostasis and increases Fe and Zn content. *Plant Physiol. Biochem.* 106, 1–10. doi: 10.1016/j.plaphy.2016.04.044
- Lin, Y. F., Liang, H. M., Yang, S. Y., Boch, A., Clemens, S., Chen, C. C., et al. (2009). *Arabidopsis* IRT3 is a zinc-regulated and plasma membrane localized zinc/iron transporter. *New Phytol.* 182, 392–404. doi: 10.1111/j.1469-8137.2009.02766.x
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Maret, W. (2004). Zinc and sulfur: a critical biological partnership. *Biochemistry* 43, 3301–3309. doi: 10.1021/bi036340p
- Mäser, P., Thomine, S., Schroeder, J. I., Ward, J. M., Hirschi, K., Sze, H., et al. (2001). Phylogenetic relationships within cation transporter families of *Arabidopsis*. *Plant Physiol.* 126, 1646–1667. doi: 10.1104/pp.126.4.1646
- Milner, M. J., Seamon, J., Craft, E., and Kochian, L. V. (2013). Transport properties of members of the ZIP family in plants and their role in Zn and Mn homeostasis. *J. Exp. Bot.* 64, 369–381. doi: 10.1093/jxb/ers315
- Mizuno, T., Hirano, K., Kato, S., and Obata, H. (2008). Cloning of ZIP family metal transporter genes from the manganese hyperaccumulator plant *Chengiopanax sciadophylloides*, and its metal transport and resistance abilities in yeast. *Soil Sci. Plant Nutr.* 54, 86–94. doi: 10.1111/j.1747-0765.2007.00206.x
- Morgounov, A., Gómez-Becerra, H., Abugalieva, A., Dzhunusova, M., Yessimbekova, M., Muminjanov, H., et al. (2007). Iron and zinc grain density in common wheat grown in Central Asia. *Euphytica* 155, 193–203. doi: 10.1007/s10681-006-9321-2
- Nakanishi, H., Ogawa, I., Ishimaru, Y., Mori, S., and Nishizawa, N. K. (2010). Iron deficiency enhances cadmium uptake and translocation mediated by the Fe²⁺ transporters OsIRT1 and OsIRT2 in rice. *Soil Sci. Plant Nutr.* 52, 464–469. doi: 10.1111/j.1747-0765.2006.00055.x
- Ouyang, X., Hong, X., Zhao, X., Zhang, W., He, X., Ma, W., et al. (2016). Knock out of the PHOSPHATE 2 gene TaPHO2-A1 improves phosphorus uptake and grain yield under low phosphorus conditions in common wheat. *Sci. Rep.* 6:29850. doi: 10.1038/srep29850
- Palmgren, M. G., Clemens, S., Williams, L. E., Krämer, U., Borg, S., Schjørring, J. K., et al. (2008). Zinc biofortification of cereals: problems and solutions. *Trends Plant Sci.* 13, 464–473. doi: 10.1016/j.tplants.2008.06.005
- Paul, T. (2015). Evaluation of methods of cotton (*Gossypium hirsutum* L.) establishment and zinc fertilization in Bt. cotton-based cropping systems. *Indian J. Agron.* 60, 372–380. doi: 10.1007/s10483-009-0301-z
- Pedás, P., Ytting, C. K., Fuglsang, A. T., Jahn, T. P., Schjørring, J. K., and Husted, S. (2008). Manganese efficiency in barley: identification and characterization of the metal ion transporter HvIRT1. *Plant Physiol.* 148, 455–466. doi: 10.1104/pp.108.118851
- Pence, N. S., Larsen, P. B., Ebbs, S. D., Letham, D. L., Lasat, M. M., Garvin, D. F., et al. (2000). The molecular physiology of heavy metal transport in the Zn/Cd hyperaccumulator *Thlaspi caerulescens*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 4956–4960. doi: 10.1073/pnas.97.9.4956
- Peng, Z. S., Li, X., Yang, Z. J., and Liao, M. L. (2011). A new reduced height gene found in the tetraploid semi-dwarf wheat landrace Aiganfanmai. *Genet. Mol. Res.* 10, 2349–2357. doi: 10.4238/2011.October.5.5
- Pinson, S., Tarpley, L., Yan, W., Yeater, K., Lahner, B., Yakubova, E., et al. (2015). World-wide genetic diversity for mineral element concentrations in rice grain. *Crop Sci.* 55, 294–311. doi: 10.2135/cropsci2013.10.0656
- Rai, V., Sanagala, R., Sinilal, B., Yadav, S., Sarkar, A. K., Dantu, P. K., et al. (2015). Iron availability affects phosphate deficiency-mediated responses, and evidence of cross-talk with auxin and zinc in *Arabidopsis*. *Plant Cell Physiol.* 56, 1107–1123. doi: 10.1093/pcp/pcv035
- Ramesh, S. A., Shin, R., Eide, D. J., and Schachtman, D. P. (2003). Differential metal selectivity and gene expression of two zinc transporters from rice. *Plant Physiol.* 133, 126–134. doi: 10.1104/pp.103.026815
- Reid, R. J., Brookes, J. D., Tester, M. A., and Smith, F. A. (1996). The mechanism of zinc uptake in plants. *Planta* 198, 39–45. doi: 10.1007/BF00197584
- Richardson, A. D., Duigan, S. P., and Berlyn, G. P. (2010). An evaluation of noninvasive methods to estimate foliar chlorophyll content. *New Phytol.* 153, 185–194. doi: 10.1104/pp.103.026815
- Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., et al. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* 18:529. doi: 10.1186/s12859-017-1934-z
- Sadeghzadeh, B. (2013). A review of zinc nutrition and plant breeding. *Soil Sci. Plant Nutr.* 13, 905–927. doi: 10.4067/S0718-95162013005000072
- Salt, D. E., Prince, R. C., Pickering, I. J., and Raskin, I. (1995). Mechanisms of cadmium mobility and accumulation in Indian Mustard. *Plant Physiol.* 109, 1427–1433. doi: 10.1104/pp.109.4.1427
- Sasaki, A., Yamaji, N., Mitani-Ueno, N., Kashino, M., and Ma, J. F. (2015). A node-localized transporter OsZIP3 is responsible for the preferential distribution of Zn to developing tissues in rice. *Plant J.* 84, 374–384. doi: 10.1111/tpj.13005
- Schneider, W. (1983). The transport and accumulation of metal ions in living organisms. *Inorg. Chim. Acta* 79:76. doi: 10.1016/S0020-1693(00)95105-3
- Sillanpää, M., and AGL (1982). *Micronutrients and the nutrient status of soils: A global study*. Rome: Rome (Italy) FAO.
- Sofo, A., Vitti, A., Nuzzaci, M., Tataranni, G., Scopa, A., Vangronsveld, J., et al. (2013). Correlation between hormonal homeostasis and morphogenic responses in *Arabidopsis thaliana* seedlings growing in a Cd/Cu/Zn multi-pollution context. *Physiol. Plant.* 149, 487–498. doi: 10.1111/pp.12050
- Tavarez, M., Macri, A., and Sankaran, R. P. (2015). Cadmium and zinc partitioning and accumulation during grain filling in two near isogenic lines of durum wheat. *Plant Physiol. Biochem.* 97, 461–469. doi: 10.1016/j.plaphy.2015.10.024
- Taylor, K. M., Morgan, H. E., Johnson, A., and Nicholson, R. I. (2004). Structure-function analysis of HKE4, a member of the new LIV-1 subfamily of zinc transporters. *Biochem. J.* 377, 131–139. doi: 10.1042/bj20031183
- Tiong, J., McDonald, G. K., Genc, Y., Pedas, P., Hayes, J. E., Toubia, J., et al. (2014). HvZIP7 mediates zinc accumulation in barley (*Hordeum vulgare*) at moderately high zinc supply. *New Phytol.* 201, 131–143. doi: 10.1111/nph.12468
- Tiong, J., McDonald, G., Genc, Y., Shirley, N., Langridge, P., and Huang, C. Y. (2015). Increased expression of six ZIP family genes by zinc (Zn) deficiency is associated with enhanced uptake and root-to-shoot translocation of Zn in barley (*Hordeum vulgare*). *New Phytol.* 207, 1097–1109. doi: 10.1111/nph.13413
- Vaucheret, H. (2006). Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev.* 20, 759–771. doi: 10.1101/gad.1410506
- Vert, G., Briat, J. F., and Curie, C. (2001). *Arabidopsis* IRT2 gene encodes a root-periphery iron transporter. *Plant J.* 26, 181–189. doi: 10.1046/j.1365-313x.2001.01018.x
- Vert, G., Grotz, N., Dédaldéchamp, F., Gaymard, F., Guerinot, M. L., Briat, J. F., et al. (2002). IRT1, an *Arabidopsis* transporter essential for iron uptake from the soil and for plant growth. *Plant Cell* 14, 1223–1233. doi: 10.1105/tpc.001388
- Welch, R. M., and Graham, R. D. (2004). Breeding for micronutrients in staple food crops from a human nutrition perspective. *J. Exp. Bot.* 55, 353–364. doi: 10.1093/jxb/erh064
- Xiao, M., Li, J., Li, W., Wang, Y., Wu, F., Xi, Y., et al. (2017). MicroRNAs activate gene transcription epigenetically as an enhancer trigger. *RNA Biol.* 14, 1326–1334. doi: 10.1080/15476286.2015.1112487
- Yamaji, N., and Ma, J. F. (2014). The node, a hub for mineral nutrient distribution in graminaceous plants. *Trends Plant Sci.* 19, 556–563. doi: 10.1016/j.tplants.2014.05.007
- Yoshida, S. (1972). Physiological aspects of grain yield. *Annu. Rev. Plant Biol.* 23, 437–464.
- Zamore, P. D., Tuschl, T., Sharp, P. A., and Bartel, D. P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at

- 21 to 23 nucleotide intervals. *Cell* 101, 25–33. doi: 10.1016/S0092-8674(00)80620-0
- Zhao, J. (2007). Interplay among nitric oxide and reactive oxygen species: a complex network determining cell survival or death. *Plant Signal. Behav.* 2, 544–547. doi: 10.4161/psb.2.6.4802
- Ziaieian, A. H., and Malakouti, M. J. (2001). “Effects of Fe, Mn, Zn and Cu fertilization on the yield and grain quality of wheat in the calcareous soils of Iran,” in *Plant Nutrition, Developments in Plant and Soil Sciences*. ed. W. J. Horst, (Dordrecht: Springer Netherlands), 840–841.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Li, Liu, Guo, Li, Nie, Chai and Zheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Identification and Validation of Stable Quantitative Trait Loci for SDS-Sedimentation Volume in Common Wheat (*Triticum aestivum* L.)

Shuai Tian¹, Minghu Zhang¹, Jinghui Li², Shaozhe Wen¹, Chan Bi¹, Huanhuan Zhao¹, Chaoyong Wei¹, Zelin Chen¹, Jiazhen Yu¹, Xintian Shi¹, Rongqi Liang¹, Chaojie Xie¹, Baoyun Li¹, Qixin Sun^{1,3}, Yufeng Zhang^{1*} and Mingshan You^{1*}

OPEN ACCESS

Edited by:

Hakan Ozkan,
Çukurova University, Turkey

Reviewed by:

Zhi Zheng,
Commonwealth Scientific
and Industrial Research Organization
(CSIRO), Australia
Reza Mir Drikvand,
Islamic Azad University, Iran

*Correspondence:

Yufeng Zhang
zhangyufeng@cau.edu.cn
Mingshan You
msyou67@cau.edu.cn

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 26 July 2021

Accepted: 08 November 2021

Published: 07 December 2021

Citation:

Tian S, Zhang M, Li J, Wen S,
Bi C, Zhao H, Wei C, Chen Z, Yu J,
Shi X, Liang R, Xie C, Li B, Sun Q,
Zhang Y and You M (2021)
Identification and Validation of Stable
Quantitative Trait Loci
for SDS-Sedimentation Volume
in Common Wheat (*Triticum*
aestivum L.).
Front. Plant Sci. 12:747775.
doi: 10.3389/fpls.2021.747775

¹ State Key Laboratory for Agrobiotechnology, Key Laboratory of Crop Heterosis and Utilization, The Ministry of Education, Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing, China, ² Wheat Center, Henan Institute of Science and Technology, Henan Provincial Key Laboratory of Hybrid Wheat, Xinxiang, China, ³ National Plant Gene Research Centre, Beijing, China

Sodium dodecyl sulfate-sedimentation volume is an important index to evaluate the gluten strength of common wheat and is closely related to baking quality. In this study, a total of 15 quantitative trait locus (QTL) for sodium dodecyl sulfate (SDS)-sedimentation volume (SSV) were identified by using a high-density genetic map including 2,474 single-nucleotide polymorphism (SNP) markers, which was constructed with a doubled haploid (DH) population derived from the cross between Non-gda3753 (ND3753) and Liangxing99 (LX99). Importantly, four environmentally stable QTLs were detected on chromosomes 1A, 2D, and 5D, respectively. Among them, the one with the largest effect was identified on chromosome 1A (designated as QSsv.cau-1A.1) explaining up to 39.67% of the phenotypic variance. Subsequently, QSsv.cau-1A.1 was dissected into two QTLs named as QSsv.cau-1A.1.1 and QSsv.cau-1A.1.2 by saturating the genetic linkage map of the chromosome 1A. Interestingly, favorable alleles of these two loci were from different parents. Due to the favorable allele of QSsv.cau-1A.1.1 was from the high-value parents ND3753 and revealed higher genetic effect, which explained 25.07% of the phenotypic variation, mapping of this locus was conducted by using BC₃F₁ and BC₃F₂ populations. By comparing the CS reference sequence, the physical interval of QSsv.cau-1A.1.1 was delimited into 14.9 Mb, with 89 putative high-confidence annotated genes. SSVs of different recombinants between QSsv.cau-1A.1.1 and QSsv.cau-1A.1 detected from DH and BC₃F₂ populations showed that these two loci had an obvious additive effect, of which the combination of two favorable loci had the high SSV, whereas recombinants with unfavorable loci had the lowest. These results provide further insight into the genetic basis of SSV and QSsv.cau-1A.1.1 will be an ideal target for positional cloning and wheat breeding programs.

Keywords: SDS-sedimentation volume, major QTL, wheat, quality, mapping

INTRODUCTION

Common wheat is one of the most widely cultivated food crops in the world and one of the important sources of carbohydrates and proteins for human beings (Osinowo, 2011). For a long time, breeders of common wheat (*Triticum aestivum* L.) have mainly committed to improving wheat yield and disease resistance (Curtis and Halford, 2014). However, high-quality wheat has been demanded by consumers and industries, and wheat quality improvement has attracted increasing attention among breeders (Guzman et al., 2016). Gluten strength has a considerable influence on the end-use quality of wheat (Rubenthaler et al., 1990; Addo et al., 1991; Slade and Levine, 1994; Kweon et al., 2011; Liu et al., 2017b), which can be measured by various tests such as sodium dodecyl sulfate (SDS)-sedimentation volume (SSV), extensograph, farinograph, alveograph, and gluten index (Huang et al., 2006; Elangovan et al., 2008; Li et al., 2009; Kerfal et al., 2010; Tsilo et al., 2011). The SSV test is well correlated with gluten strength and bread-making quality of wheat (Axford et al., 1979; He et al., 2004; Ozturk et al., 2008), and exhibits advantages such as simplicity, low cost, small sample size requirement, and high efficiency. Therefore, it has been widely used for evaluating the content and quality of gluten protein and for fast screening desired cultivars in wheat breeding programs (Clarke et al., 2000).

Some recent studies have found that the quantitative nature of SSV is closely correlated to multiple genes encoding glutenins and gliadins, such as *Glu-1*, *Glu-A3*, *Glu-B3*, and *Gli-B1* (Payne and Lawrence, 1983; Payne et al., 1984; Shewry et al., 2003; Maucher et al., 2009; Reif et al., 2011; Deng et al., 2015; Guo et al., 2020). Glutenins and gliadins are not only the most important storage proteins of wheat but also the main components of gluten protein (Gianibelli et al., 2001; Kerfal et al., 2010). Glutenins are related to the extensibility of gluten, while gliadins are associated with the elasticity of gluten (MacRitchie, 1995; Veraverbeke and Delcour, 2002; Van Der Borgh et al., 2005; Rasheed et al., 2014). The content and ratio of glutenins and gliadins are the main factors that determine the wheat processing quality (Yang et al., 2014). In addition, puroindolines are a component of wheat grain protein and are closely related to grain hardness. *Puroindoline b* (*Pinb-D1*) gene was found to be related to the variation of SSV in recent studies (Park et al., 2010, 2012; Ahn et al., 2014; Würschum et al., 2016).

Sodium dodecyl sulfate (SDS)-sedimentation volume (SSV) is a complex quantitative trait affected by both environmental and genetic factors. Quantitative trait locus (QTL) analysis is an effective approach for examining the genetic basis of quantitative traits (Doerge, 2002). Many studies have analyzed QTLs for SSV. To date, QTLs for SSV have been detected on almost all chromosomes, explained 2.2–41.4% of the phenotypic variation (Li et al., 2009; Kerfal et al., 2010; Reif et al., 2011; Deng et al., 2015; Würschum et al., 2016; Liu et al., 2017a; Mir Drikvand et al., 2018; Goel et al., 2019; Guo et al., 2020; Yang et al., 2020). However, most of the previously identified QTLs were detected only in one or two environments and could not be detected in multiple genetic backgrounds, which were not ideal targets for fine mapping and map-based cloning.

Here, to understand the genetic basis underlying SSV and provide molecular markers linked to QTL for wheat quality breeding, a doubled haploid (DH) population derived from a cross between Non-gda3753 (ND3753) and Liangxing99 (LX99) was employed to detect the QTLs associated with SSV variation. The genetic effect of two major SSV QTLs was validated.

MATERIALS AND METHODS

Plant Materials

The DH population consisting of 123 individuals was developed through *in vitro* anther culture (De Buyser and Henry, 1980) of the F₁ hybrids from a cross between ND3753 and LX99. The DH population and two parents were used for genome-wide identification of QTLs related to SSV.

For mapping of the possible QTL, ND3753 that carried the positive allele of the QTL in the confidence interval was crossed with LX99. F₁ plants were backcrossed with LX99 for three generations with insertion-deletion (InDel) marker-assisted selection to generate a BC₃F₁ population containing 418 plants. Subsequently, 126 heterozygotes lines at the QTL-anchored region were self-pollinated to BC₃F₂ containing 1,081 plants. This population with LX99 background is presented as BC₃F₂-L in the present paper.

In addition, in order to evaluate the effects of two QTL, LX99 was crossed with recurrent parent ND3753 and 64 BC₃F₁ heterozygotes lines at the QTL-anchored region were self-pollinated to construct another BC₃F₂ population containing 387 plants with marker-assisted selection. This population with the background of ND3753 is presented as BC₃F₂-N.

Field Trials

The DH population and two parents were planted in seven environments during the wheat-growing seasons of 2016, 2017, and 2018 in Beijing (BJ) (40°08'N, 116°10'E), Linfen (LF) (36°04'N, 111°31'E), Xi'an (XA) (34°16'N, 108°55'E), and Cangzhou (CZ) (38°18'N, 116°49'E), China. The seven environments, namely 2016BJ, 2017BJ, 2017LF, 2017XA, 2018BJ, 2018LF, and 2018CZ were presented in this study as E₁, E₂, E₃, E₄, E₅, E₆, and E₇, respectively. The field trials were conducted following a complete random block design with three biological replicates. However, only one biological duplication was harvested in E₇ due to an accident of field management. Then 60 seeds for each of the lines and two parents were planted in two rows of 1.5 m long and the row space was 20 cm.

The BC₃F₁, BC₃F₂-L, and BC₃F₂-N populations were all planted in Beijing. The BC₃F₁ population was planted in the wheat-growing seasons of 2018, while the BC₃F₂-L and BC₃F₂-N populations were planted in 2019. All of these backcross populations were sown in rows of 1.5 m long and 30 cm row space with a sowing density of 20 seeds per row. The BC populations were all planted in one trial and designed as a single replicate. During the whole growing season, the local standard field management methodologies were adopted for plant cultivation.

Evaluation of Traits

For DHs, 80 plants were harvested in each line of the seven environments and wheat flour of each line planted in E1 was obtained with a CD1 Quadrumat Junior laboratory mill (Chopin Technology, Paris, France), while the whole wheat flour of each line in the other six environments was produced by an XF-98B experimental mill (Zhenxing Electromechanical Instrument Factory, Cangzhou, China). SSV was determined according to a modified protocol of Axford et al. (1979) and Preston et al. (1982) using 2 g of samples. The specific procedure of the SSV test was similar to that described by Li et al. (2009). In particular, SSVs of each DH line with only one biological duplication harvesting from E7 were measured. SSV of the BC₃F₁, BC₃F₂-L, and BC₃F₂-N populations was measured with whole wheat flour from a single plant.

In addition, 300 g of grains of 30 randomly selected DH lines based on the minimum sample required for the Pearson's correlation coefficient calculated according to the formula provided by Mangard et al. (2007) and Chen et al. (2011) in E5 were ground into flour with a flour yield of approximately 60% in all samples. Their farinograph parameters (GB/ICC) were recorded by a Farinograph (DongFu JiuHeng, Beijing, China) to evaluate the correlation with SSV (Chicago, IL, United States) (ICC, 1996; Luo et al., 2018).

Genetic Map Construction

Deoxyribonucleic acid (DNA) was extracted from fresh leaves of individual DH lines and two parents using the hexadecyltrimethyl ammonium bromide (CTAB) method (Allen et al., 2006). The 15 K Axiom® Wheat Breeder single-nucleotide polymorphism (SNP) Genotyping Array (China Golden Marker Co., Beijing, China) containing 13,947 SNP markers was used to genotype the DH population and parents. SNP markers with a missing data rate > 20% were removed, and the remaining polymorphic markers were used to construct a wheat genetic map based on the inclusive composite interval mapping (ICIM) method using IciMapping v4.1 (Chinese Academy of Agricultural Sciences, China) and MapChart v2.32 (Plant Research International, P.O. Box 16, 6700 AA Wageningen) (Voorrips, 2002). The physical locations of unique SNP markers were obtained from the International Wheat Genome Sequencing Consortium (Appels et al., 2018).

Quantitative Trait Locus Mapping

The average value of SSV in each environment and the BLUP were employed for QTL analysis using inclusive composite interval mapping (ICIM¹) method in software IciMapping v4.1 (Meng et al., 2015). A QTL with LOD $\geq$ 2.5 was defined as a significant QTL. The confidence intervals ($\pm$ 2 LOD away from the peaks of likelihood ratios) of several QTLs were coincident, which were preliminarily considered as the same QTL. In this study, the QTL that can be detected in three or more environments is defined as an environmentally stable QTL.

¹<http://www.isbreeding.net/software/?type=detail&id%20=18>

Re-sequencing and InDel Markers Development

High-quality genomic DNA of ND3753 and LX99 was extracted to construct paired-end-sequencing libraries. According to the procedures described by Li et al. (2020), the parents were re-sequenced with an average sequencing depth of $6 \times$ and paired-end reads of length 150 bp for two parents using the Illumina HiSeq X Ten platform (Illumina, California, United States), and the re-sequencing data were processed. The InDels were identified using the HaplotypeCaller module of the Genome Analysis Toolkit (GATK). The InDel markers were developed based on the sequence difference between the parents around the target region. Primer3 version 0.4.0² was used to design the sequences of InDel primers.

Deoxyribonucleic acid (DNA) amplification was programmed for an initial 5 min at 94°C, then followed by 35 cycles of 30 s at 94°C, 30 s at 56°C, and 30 s at 72°C, and finally 5 min at 72°C. A 10 μ L PCR reaction system was used, containing 5 μ L of $2 \times$ Taq PCR StarMix (GenStar, Beijing, China) (for PAGE), 1.5 μ L of DNA template (about 50–100 ng), 1.5 μ L of each InDel primer, and double-distilled H₂O. The PCR products were analyzed on 8% non-denaturing polyacrylamide gels with silver staining.

Validation and Mapping and Annotation of Putative Genes

Insertion-deletion (InDel) markers tightly linked to *QSSv.cau-1A.1.1* and *QSSv.cau-1A.1.2* were used to genotype DH, BC₃F₁, BC₃F₂-L, and BC₃F₂-N populations. The putatively annotated high confidence (HC) genes located between the flanking markers of *QSSv.cau-1A.1.1* were acquired based on the reference genome of *T. aestivum* cv. Chinese Spring³.

Statistical Analysis

To conduct phenotypic statistical analysis and calculate correlation coefficients between SSV and farinograph parameters, IBM SPSS Statistics 21.0 (SPSS, Chicago, United States) was used. R software v3.6.2⁴ was used to perform the Shapiro-Wilk test across seven environments and the best linear unbiased prediction (BLUP), as well as to estimate the broad-sense heritability (h^2_B) following the formula: $h^2_B = \sigma_g^2 / (\sigma_g^2 + \sigma_{ge}^2/n + \sigma^2/nr)$, as described by Liu et al. (2014).

RESULTS

Phenotypic Analysis

Descriptive statistics for SSV of two parents and the DH population in the seven environments are shown in **Table 1**. SSV of ND3753 was significantly higher than that of LX99 in all environments. SSV showed bi-directional transgressive

²<http://bioinfo.ut.ee/primer3-0.4.0/>

³https://urgi.versailles.inra.fr/download/iwgs/IWGS_RefSeq_Annotations/v1.0/

⁴<https://www.r-project.org/>

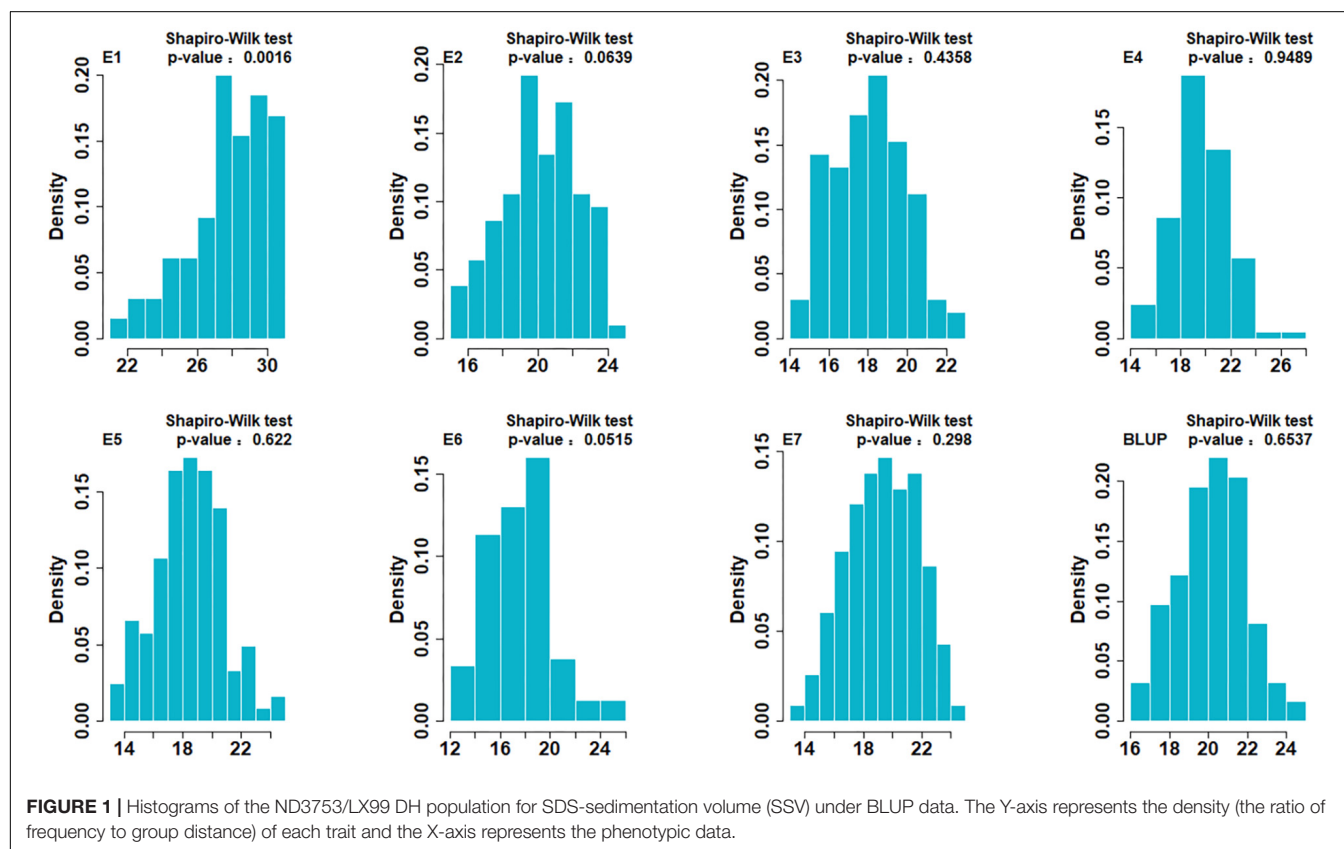
TABLE 1 | Descriptive statistics of two parents and DH population for SDS-sedimentation volume (SSV) under seven environments.

Trait	Environment ^a	Parents		DH population			
		ND3753	LX99	Range	Mean	SD ^b	<i>h</i> ² B ^c
SSV (mL)	E1	23.8 ± 0.9	20.0 ± 1.6	15.6–26.3	21.6	2.5	0.86
	E2	23.4 ± 2.6	19.4 ± 0.5	15.3–25.8	21.2	2.3	0.91
	E3	20.6 ± 0.8	15.8 ± 0.4	13.2–24.7	18.2	2.1	0.80
	E4	23.1 ± 1.4	16.4 ± 1.9	15.0–26.1	19.8	2.2	0.88
	E5	20.9 ± 3.3	17.2 ± 1.1	13.6–24.5	18.5	2.3	0.86
	E6	21.0 ± 0.4	14.2 ± 0.8	12.0–25.3	17.5	2.5	0.86
	E7	22.4	17.0	13.7–24.4	19.3	2.4	–

^aE1, 2016–2017 (Beijing); E2, 2017–2018 (Beijing); E3, 2017–2018 (Linfen); E4, 2017–2018 (Xi'an); E5, 2018–2019 (Beijing); E6, 2018–2019 (Linfen); E7, 2018–2019 (Cangzhou).

^bSD, standard deviation.

^c*h*² B, Broad sense heritability based on a family mean basis was estimated under individual environments.



segregation, suggesting that both parents have increasing alleles for SSV. The broad-sense heritability values in all environments were greater than 0.8, indicating that SSV was mainly controlled by genetic factors. The result of the Shapiro-Wilk test displayed that SSV exhibited normal distribution under six environments and BLUP value, indicating SSV was determined by many genes (Figure 1). Pearson's correlation coefficient analysis was carried out between farinograph parameters and SSVs of 30 DH lines planted in E5 which indicated a significant positive correlation between the two (Table 2).

Linkage Map Construction

A total of 2,523 SNP markers showed polymorphisms between the two parents. Finally, 2,474 SNP markers participated in the map construction and were mapped to 21 linkage groups, covering the 21 chromosomes of common wheat (Supplementary Table 1). The total length of the map was 7,349.01 cm, and the average interval distance between two adjacent markers was 7.24 cm (Supplementary Table 1). The A genome contained the most SNP markers (963), followed by the B genome (902), while the D genome had the least (609) (Supplementary Table 1). The total length of chromosome 7D

TABLE 2 | Pearson's correlation analysis among SDS-sedimentation volume (SSV), dough stability time (DST), dough developing time (DDT), and water absorption (WA) in E5/2018–2019 (Beijing).

Trait	SSV	DST	DDT	WA
SSV	1			
DST	0.614**	1		
DDT	0.595**	0.918**	1	
WA	0.536**	0.103	0.234	1

**Correlation is significant at the 0.01 level (2-tailed).

was the largest (619.02 cm), while that of chromosome 4B was the smallest (180.86 cm) (**Supplementary Table 1**). Chromosome 5A harbored the most SNP markers (236), while chromosome 6D contained the least (42) (**Supplementary Table 1**).

Quantitative Trait Locus Analysis

A total of 15 QTLs were detected on 11 chromosomes (1A, 1B, 1D, 2A, 2D, 4B, 4D, 5A, 5D, 6B, and 6D) in the seven environments (**Table 3** and **Supplementary Table 2**). Four environmentally stable QTLs (*Q_{SSV.cau-1A.1}*, *Q_{SSV.cau-1A.2}*, *Q_{SSV.cau-2D}* and *Q_{SSV.cau-5D.1}*) were identified on chromosomes 1A, 1A, 2D, and 5D, respectively (**Table 3**). The favorable allele of *Q_{SSV.cau-2D}* came from LX99, while the superior alleles of the other three QTLs were contributed by ND3753. The major QTL *Q_{SSV.cau-1A.1}* was repeatedly detected in five environments and the BLUP data, explaining 39.67% of the phenotypic variation in the BLUP analysis (**Table 3**). *Q_{SSV.cau-2D}* and *Q_{SSV.cau-5D.1}* contributed 3.17 and 4.82% of the phenotypic variation in the BLUP analysis, respectively. *Q_{SSV.cau-1A.2}* explained 8.17–18.62% of the

phenotypic variation. The remaining 11 were putative QTLs (**Supplementary Table 2**).

Verification of *Q_{SSV.cau-1A.1}*

Considering its stability and the genetic effect, *Q_{SSV.cau-1A.1}* was chosen to saturate the positioning interval. Then 15 InDel markers near this region were developed according to the re-sequencing results of two parents (**Table 3** and **Supplementary Table 3**) and a new genetic linkage map of 1A long arm was constructed. QTL remapping detected an additional QTL located next to the original interval. These two QTLs were named *Q_{SSV.cau-1A.1.1}* and *Q_{SSV.cau-1A.1.2}* (**Figure 2**).

Q_{SSV.cau-1A.1.1} was detected in six environments and BLUP, explaining 17.21–26.47% of the phenotypic variation, and the favorable allele was from ND3753. The confidence interval was between the markers *1A248* and *1A332* corresponding CS physical position of 386,222,224–463,923,653 bp (Appels et al., 2018; **Figure 2** and **Supplementary Table 5**). *Q_{SSV.cau-1A.1.2}* was repeatedly detected in three environments as well as BLUP data, contributing 7.02–12.13% of the phenotypic variation and LX99 contributed the favorable allele. The physical position of *Q_{SSV.cau-1A.1.2}* located on 462,634,655–492,004,197 bp by comparing flanking markers *1A36* and *1A215* to CS RefSeqv1.0 (Appels et al., 2018; **Figure 2** and **Supplementary Table 5**).

Effects of *Q_{SSV.cau-1A.1.1}* and *Q_{SSV.cau-1A.1.2}* in Different Genetic Backgrounds

The flanking markers *1A248*, *1A332* and *1A36*, *1A215* delimiting confidence intervals of *Q_{SSV.cau-1A.1.1}* and *Q_{SSV.cau-1A.1.2}* (**Figure 2**) separately were used to detect genotypes in the DH

TABLE 3 | The QTL regions harboring environmentally stable QTLs for SSV in the ND3753/LX99 DH population.

QTL	Environment	Flanking marker	Position (cM)	Interval (cM)	LOD	PVE ^a (%)	Additive ^b
<i>Q_{SSV.cau-1A.1}</i>	E3	AX-109863151 and AX-110089093	164.0	163.1–166.1	6.15	13.69	0.83
	E4	AX-109863151 and AX-110089093	164.0	163.1–166.1	4.15	15.94	0.81
	E5	AX-109863151 and AX-110089093	164.0	163.1–166.1	7.40	19.70	0.88
	E6	AX-109863151 and AX-110089093	164.0	163.1–166.1	3.37	9.78	0.74
	E7	AX-109863151 and AX-110089093	164.0	163.1–166.1	4.99	12.14	0.77
	BLUP ^c	AX-109863151 and AX-110089093	164.0	163.1–165.1	23.65	39.67	1.32
<i>Q_{SSV.cau-1A.2}</i>	E1	AX-109863129 and AX-111450961	220.0	211.1–229.1	3.94	18.62	0.92
	E2	AX-109863129 and AX-111450961	218.0	213.1–224.1	8.42	21.03	1.01
	E5	AX-110673287 and AX-111688135	231.0	223.1–232.1	3.43	8.17	0.56
<i>Q_{SSV.cau-2D}</i>	E2	AX-110872666 and AX-110773527	143.0	142.9–143.9	3.19	7.15	–0.57
	E5	AX-111430851 and AX-110773527	144.0	143.9–148.9	4.87	12.20	–0.68
	E7	AX-110773527 and AX-109246010	153.0	151.9–155.9	5.79	13.75	–0.81
	E6	AX-110773527 and AX-109246010	155.0	151.9–155.9	4.74	13.56	–0.86
	BLUP	AX-110872666 and AX-110773527	143.0	142.9–143.9	2.91	3.17	–0.37
<i>Q_{SSV.cau-5D.1}</i>	E2	AX-89753391 and AX-109174882	501.0	495.9–506.9	7.30	18.31	0.92
	E4	AX-89753391 and AX-109174882	502.0	494.9–505.9	3.48	12.91	0.74
	E7	AX-89753391 and AX-109174882	501.0	490.9–509.9	3.18	7.40	0.60
	BLUP	AX-89753391 and AX-109174882	502.0	493.9–505.9	4.32	4.82	0.46

^aPVE, Phenotypic variation explained by the stable QTLs.

^bAdditive, the additive effect of a QTL, positive values: a favorable allele from ND3753; negative values: a favorable allele from LX99.

^cBLUP, phenotype values based on best linear unbiased prediction.

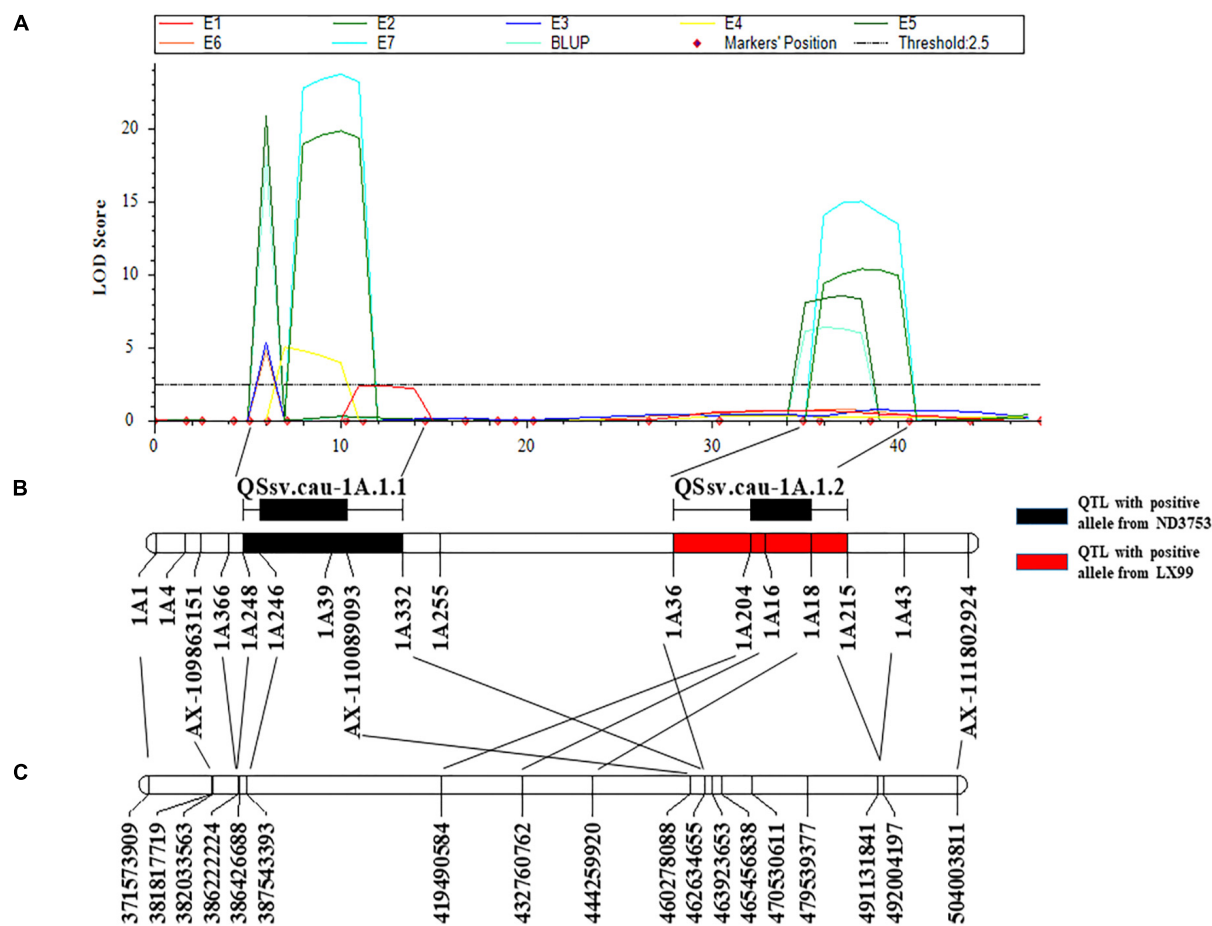


FIGURE 2 | Genetic and physical locations of *QSsv.cau-1A.1.1* and *QSsv.cau-1A.1.2*. **(A)** QTL mapping for SSV in seven individual environments (E1–E7) and BLUP using a saturated genetic map of chromosome arm 1AL. **(B)** Saturated genetic map of chromosome arm 1AL with newly integrated InDel markers in the DH population. The black and red rectangles indicate QTLs with positive alleles from the parent ND3753 and the parent LX99, respectively. The black horizontal lines and bars above the genetic map represent the confidence interval of the two QTLs. **(C)** Corresponding physical positions according to the Chinese Spring IWGSC RefSeq v1.0 sequence.

and BC_3F_2 -N populations. AA and aa represented genotypes with homozygous favorable and unfavorable alleles of *QSsv.cau-1A.1.1* from ND3753, respectively, whereas BB and bb were symbols of that of *QSsv.cau-1A.1.2* from LX99.

DH lines could be grouped into three genotypes which included two parental genotypes AAbb, aaBB, and one recombined genotype aabb, and each contained 51, 59, and 13 lines. The reason for the absence of genotype AABB remained unknown. The average SSV values in BLUP of genotype AAbb, aaBB, and aabb were 21.2, 19.9, and 17.8 mL, respectively, of which AAbb was significantly higher than aaBB and both were significantly higher than that of aabb (Figure 3A). This suggested that *QSsv.cau-1A.1.1* had a stronger effect on SSV than *QSsv.cau-1A.1.2*, which was consistent with their contribution rates of phenotypic variation and additive effects in QTL analysis.

In BC_3F_2 -N population four allele combinations AaBB, AAbb, aaBB, and aabb were identified, which had SSV average values of 20.4, 17.5, 17.9, and 15.8 mL from 2, 23, 20, and 19 plants, respectively. Similarly, the SSV values of genotypes AaBB, AAbb,

and aaBB were significantly higher than that of aabb (Figure 3B). However, although genotype AaBB had the distinct highest value, three genotypes with favorable alleles had no significant difference. This was possibly due to the too-small sample number of genotypes AaBB. Nevertheless, all results above could still prove that the favorable allele has positive effects.

This study did not find the combination type of AABB but found the type AaBB in the BC_3F_2 -N population. In summary, the combination of two favorable loci had the high SSV, whereas recombinants with unfavorable loci had the lowest. In future research, the homozygous lines of the BC_3F_2 -N population will be extracted to develop the corresponding NIL pairs to further verify their effects and examine their genetic effect on some wheat qualities, such as gluten content, extensograph, farinograph, alveograph, and gluten index.

Mapping of *QSsv.cau-1A.1.1*

In order to further verify and narrow down the confidence interval of *QSsv.cau-1A.1.1*, the BC_3F_1 population in the

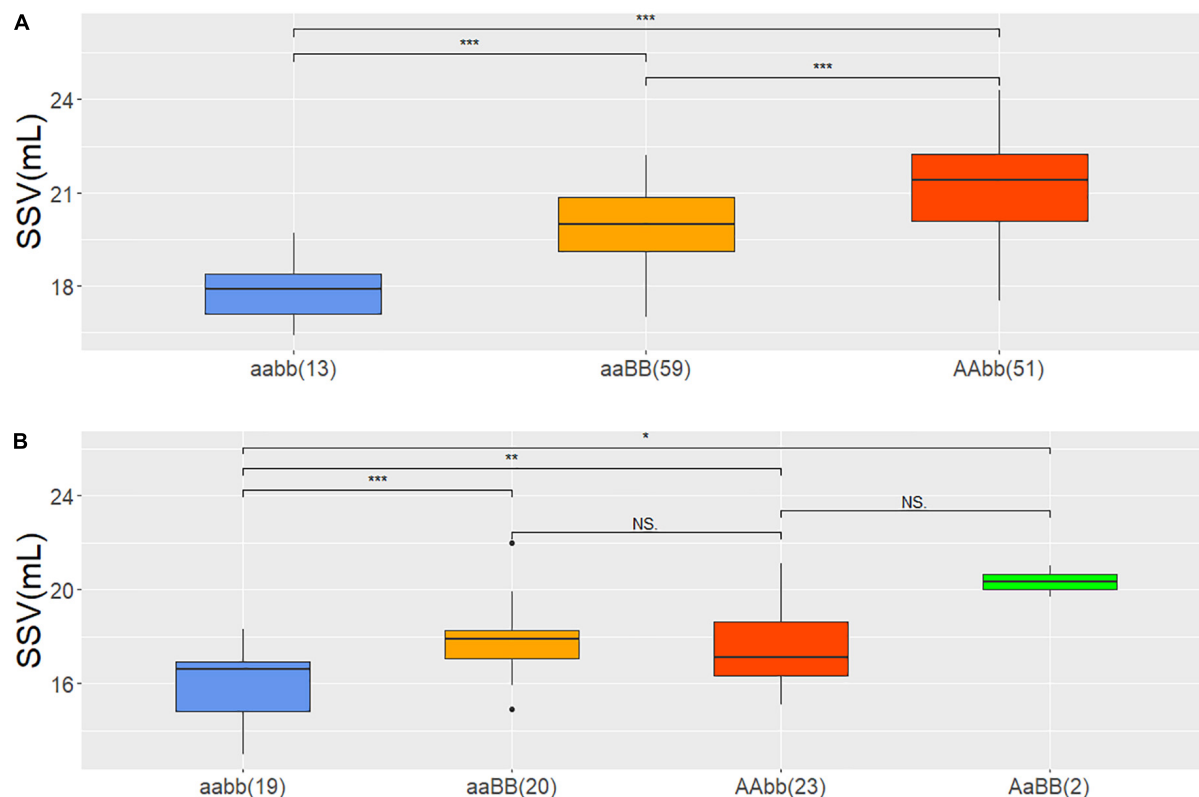


FIGURE 3 | The phenotypic effect of *QsSv.cau-1A.1.1* and *QsSv.cau-1A.1.2* in the DH and BC₃F₂-N populations according to the BLUP value for SSV and the means of SSVs of different types, respectively. **(A)** DH population; **(B)** BC₃F₂-N population; *, **, and *** indicate significant differences at the 0.05, 0.01, and 0.001 levels, respectively (Student's *t*-test). aabb: *QsSv.cau-1A.1.1* carrying homozygous alleles from LX99 and *QsSv.cau-1A.1.2* carrying homozygous alleles from ND3753; aaBB: *QsSv.cau-1A.1.1* carrying homozygous alleles from LX99 and *QsSv.cau-1A.1.2* carrying homozygous alleles from LX99; AAbb: *QsSv.cau-1A.1.1* carrying homozygous alleles from ND3753 and *QsSv.cau-1A.1.2* carrying homozygous alleles from ND3753; AaBB: *QsSv.cau-1A.1.1* carrying heterozygous alleles and *QsSv.cau-1A.1.2* carrying homozygous alleles from LX99. The numbers in parentheses indicate sample size.

background of LX99 was genotyped using six InDel markers, and four recombinant types were obtained (Figure 4B). The SSV values of types 3 and 4 were similar and were significantly lower than that of types 1 and 2, indicating that *QsSv.cau-1A.1.1* was delimited to the interval between markers *1A1* and *1A366*.

The BC₃F₁ individuals with heterozygous genotypes between markers *1A1* and *1A39* were selected to generate the BC₃F₂-L population. The six markers between *1A1* and *1A39* were used to genotype the BC₃F₂-L population, and six recombinant types were obtained (Figure 4C). The SSV value of type 6 was similar to that of types 4 and 5 but was significantly lower than that of the other types (Figure 4D). These results once again proved the effectiveness of *QsSv.cau-1A.1.1* and further confirmed that its location was between markers *1A1* and *1A366* (Figure 4C). There was no significant phenotypic difference between the heterozygous genotype and the ND3753 genotype, which implied that the ND3753 genotype was dominant. The corresponding physical interval of CS between these two markers was 14.9 Mb, with 89 putative high-confidence annotated genes (Supplementary Table 6). In addition, further fine mapping of *QsSv.cau-1A.1.1* is under research.

DISCUSSION

Correlation of Sodium Dodecyl Sulfate-Sedimentation Volume and Farinograph Parameters

Sodium dodecyl sulfate (SDS)-sedimentation volume (SSV) is a comprehensive indicator for indirectly testing wheat quality and one of the important tests to evaluate the gluten strength of flour and is closely related to the processing and baking quality of flour (Axford et al., 1978; Peña-Bautista, 2002; He et al., 2004). SSV is well correlated with other quality traits, such as grain protein content, gluten index, wet gluten content, bread volume, and farinograph parameters (Cubadda et al., 1992).

Our results showed that SSV was significantly positively correlated with stability time, formation time, and water absorption of dough, which is consistent with previous studies (Table 2; Cubadda et al., 1992). This confirms that SSV can be used as a suitable substitute for farinograph indicators that require a great number of samples to evaluate the rheological properties of wheat dough, thereby indirectly measuring the baking and processing quality of wheat flour.

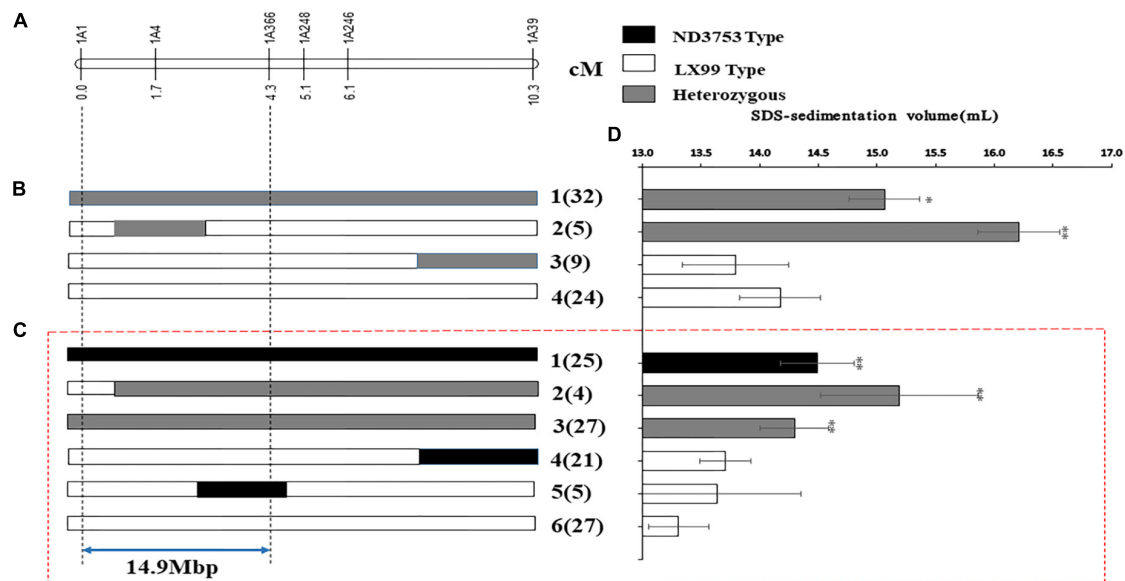


FIGURE 4 | Mapping of *QSsv-cau-1A.1.1*. **(A)** Genetic location of the region of interest on chromosome arm 1AL. Graphical illustration of recombinant genotypes from **(B)** the BC₃F₁ population and **(C)** the BC₃F₂ population in the LX99 background. **(D)** SSV values (mean $\pm$ SD). Black, gray, and white bars represent the ND3753 genotype, heterozygous genotype, and the LX99 genotype, respectively. The arrow indicates the 14.9-Mb mapping interval. Significant differences by comparing with the ND3753 genotype are indicated by * ($p < 0.05$), and ** ($p < 0.01$) (Student's *t*-test). The numbers in and outside the parentheses indicate sample size and recombinant types, respectively.

Novel Quantitative Trait Locus for Sodium Dodecyl Sulfate-Sedimentation Volume

We compared the physical locations of QTLs for SSV reported in previous studies and those revealed in this study (Li et al., 2009; Reif et al., 2011; Deng et al., 2015; Würschum et al., 2016; Yang et al., 2020). The physical position of *QSsv-cau-1A.2* was agreed with that of a previously reported QTL controlling SSV (Yang et al., 2020). SSV was found to be affected by allelic variations at *Glu-A1* (508,726,618–508,725,448 bp, RefSeqv1.0) and *Glu-A3* (4,203,001–4,202,275 bp, RefSeqv1.0) loci in several previous studies (Li et al., 2009; Reif et al., 2011; Deng et al., 2015; Würschum et al., 2016). Some QTLs associated with SSV was reported on chromosome 1A. For instance, Yang et al. (2020) identified a QTL (540,660,000–544,610,000 bp, RefSeqv1.0) for SSV that is located on chromosome 1A by genome-wide association study (GWAS). However, the physical position of *QSsv-cau-1A.1.1* (371,573,909–386,426,688 bp, RefSeqv1.0) and *QSsv-cau-1A.1.2* (419,490,584–492,004,197 bp, RefSeqv1.0) did not overlap with those of the above-mentioned QTLs/genes, suggesting these two QTLs may be novel. Li et al. (2009) identified a QTL (470,230,000–570,420,000 bp, RefSeqv1.0) for SSV on chromosome 2D using a recombinant inbred line population. Four QTLs (16,340,000, 59,102,000, 615,470,000, 646,600,000 bp, RefSeqv1.0) controlling SSV were reported to be located on chromosome 2D by multi-locus GWAS (Yang et al., 2020). However, the physical position of these QTLs and that of *QSsv-cau-2D* (140,759,212–467,689,413 bp, RefSeqv1.0) were not consistent, indicating that *QSsv-cau-2D* may also

be a novel QTL. Li et al. (2009) identified a QTL for SSV on chromosome 5DS. SSV was found to be affected by the allelic variation at the *Pinb-D1* locus on chromosome 5DS in some previous studies (Li et al., 2009; Reif et al., 2011; Deng et al., 2015; Würschum et al., 2016). However, the physical locations of these QTLs/genes and *QSsv-cau-5D.1* does not match, implying that *QSsv-cau-5D.1* on chromosome 5DL may be a novel QTL.

Sodium dodecyl sulfate (SDS)-sedimentation volume (SSV) is a quantitative trait affected by both environmental and genetic factors; thus, some QTLs can only be detected in specific environments (**Supplementary Table 2**). We found 11 such QTLs located on chromosomes 1B, 1D, 2A, 4B, 4D, 5A, 5D, 6B, and 6D, which is consistent with previous results (Li et al., 2009; Kerfal et al., 2010; Reif et al., 2011; Deng et al., 2015; Würschum et al., 2016; Liu et al., 2017a; Mir Drikvand et al., 2018; Goel et al., 2019; Guo et al., 2020; Yang et al., 2020). The *Glu-D1* gene, which is located in the interval of *QSsv-cau-1D*, may be a candidate gene for *QSsv-cau-1D*. *QSsv-cau-4B.1* and *QSsv-cau-4D* was located next to the dwarf genes *Rht-B1* and *Rht-D1*, respectively. Previous studies have also revealed QTL-enrichment areas near *Rht-B1* and *Rht-D1*, which are associated with kernel size, kernel hardness, kernel protein, pasting properties, and mixing properties (Shanhong et al., 2001; Li et al., 2006; Wang et al., 2012, 2017; Patil et al., 2013; Zhang et al., 2013; Jin et al., 2016; Liu et al., 2017b).

However, there is not enough evidence to support the correlation between *Rht-B1/D1* and SSV, and the gene that controls SSV near *Rht-B1* and *Rht-D1* has not been cloned. We hypothesized: (1) there may be other genes affecting quality traits

near *Rht-B1* and *Rht-D1*; (2) allelic variations between *Rht-B1* and *Rht-D1* may also regulate certain quality traits, such as SSV. However, these hypotheses need to be further tested. The co-localization of dwarf genes and QTLs related to quality traits may also remind breeders to consider the selection of plant height and grain quality in the wheat breeding program.

Genetic Effects and Putative Annotated Genes of the Major Quantitative Trait Locus

In recent years, a large number of QTLs for SSV have been identified and characterized through GWAS and linkage analysis, and some of these QTLs are related to allelic variants of *Glu-1*, *Glu-A3*, *Glu-B3*, *Gli-B1*, and *Pina-D1* (Ahn et al., 2014; Deng et al., 2015; Guo et al., 2020). However, most of the other QTLs have not been further verified or fine mapped. SSV is a typical quantitative trait with a complex genetic mechanism. The lack of information on the authenticity and genetic effects of these QTLs for SSV not only hinders the exploration of their genetic and molecular mechanisms but also fails to provide breeders with sufficient new high-quality genetic resources for wheat quality improvement. In this study, we verified the effects of *QSsv.cau-1A.1.1* and *QSsv.cau-1A.1.2* on SSV and the interaction between the two QTLs in the DH and BC3F2-N populations. Therefore, the InDel markers are closely linked to *QSsv.cau-1A.1.1* and *QSsv.cau-1A.1.2* developed in this study can be used by breeders to aggregate high-quality genes for wheat quality improvement.

In particular, *QSsv.cau-1A.1.1* was delimited to an approximate 14.9 Mb between markers *1A1* and *1A366* (Figure 4). We conduct an orthologous analysis for the candidate region to predict HC genes in *QSsv.cau-1A.1.1*, but no ones are associated with SSV in *Oryza sativa* and *Arabidopsis thaliana* (Supplementary Table 6). This may be because *Oryza sativa* and *Arabidopsis thaliana* do not have gluten, and SSV is related to gluten strength in the common wheat.

REFERENCES

- Addo, K., Pomeranz, Y., Huang, M. L., Rubenthaler, G. L., and Jeffers, H. C. (1991). Steamed Bread. II. Role of protein content and strength. *Cereal Chem.* 68, 39–42.
- Ahn, J. H., Kang, C. S., Jeung, J. U., Baik, B. K., and Park, C. S. (2014). Effect of allelic variations at the *Glu-D1*, *Glu-A3*, *Glu-B3* and *Pinb-D1* loci on flour characteristics and bread loaf volume. *Int. Food Res. J.* 21, 1141–1149.
- Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S., and Thompson, W. F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* 1, 2320–2325. doi: 10.1038/nprot.2006.384
- Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191
- Axford, D., McDermott, E. E., and Redman, D. G. (1979). Note on sodium dodecyl sulfate test of breadmaking quality; Comparison with Pelshenke and Zeleny test. *Cereal Chem.* 56, 582–584.
- Axford, D. W. E., McDermott, E. E., and Redman, D. G. (1978). Small scale tests of bread making quality. *Milling Feed Fert.* 13, 18–20.

DATA AVAILABILITY STATEMENT

The raw data of Nongda3753 and Liangxing99 presented in the study are deposited in the NCBI Sequence Read Archive repository, accession number PRJNA722149.

AUTHOR CONTRIBUTIONS

MY conceived the project. ST performed the research, constructed the linkage map, developed an InDel marker of the QTL region of interest, and developed the BC₃F₁ and BC₃F₂ populations. JL, MZ, and SW participated in the field trials. ST, CB, HZ, CW, ZC, JY, and XS performed the phenotypic analysis. YZ, CX, BL, RL, and QS assisted in revising the manuscript. ST and YZ analyzed the experimental results. ST and MY wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the National Key Research and Development Program of China (Grant No. 2016YFD0101602).

ACKNOWLEDGMENTS

The authors would like to thank Mingyi Zhang (Shanxi Academy of Agricultural Sciences) for help with field trials.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.747775/full#supplementary-material>

- Chen, K., Zhang, P., Deng, G., and Zou, C. (2011). Electromagnetic correlation attacks and the minimum number of attack traces analyzing. *J. Huazhong Univ. Sci. Technol.* 39:1009.
- Clarke, J. M., Clarke, F. R., Ames, N. P., McCaig, T. N., and Knox, R. E. (2000). Evaluation of predictors of quality for use in early generation selection. *Options Méditer. Sér. A Sémin. Méditerranéens* 40, 439–446.
- Cubadda, R., Carcea, M., and Pasqui, L. A. (1992). Suitability of the Gluten Index method for assessing gluten strength in durum-wheat and semolina. *Cereal Foods World* 37, 866–869.
- Curtis, T., and Halford, N. G. (2014). Food security: the challenge of increasing wheat yield and the importance of not compromising food safety. *Ann. Appl. Biol.* 164, 354–372. doi: 10.1111/aab.12108
- De Buyser, J., and Henry, Y. (1980). Induction of haploid and diploid plants through in vitro anther culture of haploid wheat (n=3x=21). *Theor. Appl. Genet.* 57, 57–58. doi: 10.1007/BF00745029
- Deng, Z., Tian, J., Chen, F., Li, W., Zheng, F., Chen, J., et al. (2015). Genetic dissection on wheat flour quality traits in two related populations. *Euphytica* 203, 221–235. doi: 10.1007/s10681-014-1318-7
- Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet.* 3, 43–52. doi: 10.1038/nrg703

- Elangovan, M., Rai, R., Dholakia, B. B., Lagu, M. D., Tiwari, R., Gupta, R. K., et al. (2008). Molecular genetic mapping of quantitative trait loci associated with loaf volume in hexaploid wheat (*Triticum aestivum*). *J. Cereal Sci.* 47, 587–598. doi: 10.1016/j.jcs.2007.07.003
- Gianibelli, M. C., Larroque, O. R., MacRitchie, F., and Wrigley, C. W. (2001). Biochemical, genetic, and molecular characterization of wheat glutenin and its component subunits. *Cereal Chem.* 78, 635–646. doi: 10.1094/CCHEM.2001.78.6.635
- Goel, S., Singh, K., Singh, B., Grewal, S., Dwivedi, N., Alqarawi, A. A., et al. (2019). Analysis of genetic control and QTL mapping of essential wheat grain quality traits in a recombinant inbred population. *PLoS One* 14:e0200669. doi: 10.1371/journal.pone.0200669
- Guo, Y., Zhang, G., Guo, B., Qu, C., Zhang, M., Kong, F., et al. (2020). QTL mapping for quality traits using a high-density genetic map of wheat. *PLoS One* 15:e0230601. doi: 10.1371/journal.pone.0230601
- Guzman, C., Peña, R. J., Singh, R., Autrique, E., Dreisigacker, S., Crossa, J., et al. (2016). Wheat quality improvement at CIMMYT and the use of genomic selection on it. *Appl. Transl. Genom.* 11, 3–8. doi: 10.1016/j.atg.2016.10.004
- He, Z. H., Yang, J., Zhang, Y., Quail, K. J., and Peña, R. J. (2004). Pan bread and dry white Chinese noodle quality in Chinese winter wheats. *Euphytica* 139, 257–267. doi: 10.1007/s10681-004-3283-z
- Huang, X. Q., Cloutier, S., Lycar, L., Radovanovic, N., Humphreys, D. G., Noll, J. S., et al. (2006). Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). *Theor. Appl. Genet.* 113, 753–766. doi: 10.1007/s00122-006-0346-7
- ICC (1996). *Standard Methods of The International Association for Cereal Science and Technology*. Vienna: ICC.
- Jin, H., Wen, W., Liu, J., Zhai, S., Zhang, Y., Yan, J., et al. (2016). Genome-wide QTL mapping for wheat processing quality parameters in a Gaocheng 8901/Zhoumai 16 recombinant inbred line population. *Front. Plant Sci.* 7:1032. doi: 10.3389/fpls.2016.01032
- Kerfal, S., Giraldo, P., Rodriguez-Quijano, M., Vázquez, J. F., Adams, K., Lukow, O. M., et al. (2010). Mapping quantitative trait loci (QTLs) associated with dough quality in a soft-hard bread wheat progeny. *J. Cereal Sci.* 52, 46–52. doi: 10.1016/j.jcs.2010.03.001
- Kweon, M., Slade, L., and Levine, H. (2011). Solvent retention capacity (SRC) testing of wheat flour: principles and value in predicting flour functionality in different wheat-based food processes and in wheat breeding-A review. *Cereal Chem.* 88, 537–552. doi: 10.1094/CCHEM-07-11-0092
- Li, J., Wen, S., Fan, C., Zhang, M., Tian, S., Kang, W., et al. (2020). Characterization of a major quantitative trait locus on the short arm of chromosome 4B for spike number per unit area in common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 133, 2259–2269. doi: 10.1007/s00122-020-03595-z
- Li, X. P., Lan, S. Q., Liu, Y. P., Gale, M. D., and Worland, T. J. (2006). Effects of different Rht-B1b, Rht-D1b and Rht-B1c dwarfing genes on agronomic characteristics in wheat. *Cereal Res. Commun.* 34, 919–924. doi: 10.1556/CRC.34.2006.2-3.220
- Li, Y., Song, Y., Zhou, R., Branlard, G., and Jia, J. (2009). Detection of QTLs for bread-making quality in wheat using a recombinant inbred line population. *Plant Breed.* 128, 235–243. doi: 10.1111/j.1439-0523.2008.01578.x
- Liu, G., Jia, L., Lu, L., Qin, D., Zhang, J., Guan, P., et al. (2014). Mapping QTLs of yield-related traits using RIL population derived from common wheat and Tibetan semi-wild wheat. *Theor. Appl. Genet.* 127, 2415–2432. doi: 10.1007/s00122-014-2387-7
- Liu, T., An, Y., Liu, K., Wang, F., Xie, C., Zhang, Y., et al. (2017a). A genetic analysis of the quality of northern-style Chinese steamed bread. *Mol. Breed.* 37:41. doi: 10.1007/s11032-016-0593-5
- Liu, T., Liu, K., Wang, F., Zhang, Y., Li, Q., Zhang, K. R., et al. (2017b). Conditional and unconditional QTLs mapping of gluten strength in common wheat (*Triticum aestivum* L.). *J. Integr. Agric.* 16, 2145–2155. doi: 10.1016/S2095-3119(16)61564-2
- Luo, D., Kou, X., Zhang, T., Nie, Y., Xu, B., Li, P., et al. (2018). Effect of inulin on rheological properties of soft and strong wheat dough. *Int. J. Food Sci. Technol.* 53, 1648–1656. doi: 10.1111/ijfs.13748
- MacRitchie, F. (1995). “Role of polymeric proteins in flour functionality,” in *Proceedings of the International Symposium, Wheat Kernel Proteins: Molecular and Functional Aspects*, Viterbo, 145–150.
- Mangard, S., Oswald, E., and Popp, T. (2007). *Power Analysis Attacks: Revealing the Secrets of Smart Cards*. Berlin: Springer.
- Maucher, T., Figueroa, J. D. C., Reule, W., and Peña, R. J. (2009). Influence of low molecular weight glutenins on viscoelastic properties of intact wheat kernels and their relation to functional properties of Wheat Dough. *Cereal Chem.* 86, 372–375. doi: 10.1094/CCHEM-86-4-0372
- Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001
- Mir Drikvand, R., Najafian, G., Bihamta, M. R., and Ebrahimi, A. (2018). Mapping some seed quality traits in bread wheat (*Triticum aestivum* L.) by association mapping using SSR markers. *J. Appl. Biotechnol. Rep.* 5, 92–99. doi: 10.29252/JABR.05.03.02
- Osinowo, F. (2011). *Food and Agricultural Organization of the United Nations, State of the World's Forests*.
- Ozturk, S., Kahraman, K., Tiftik, B., and Koxsel, H. (2008). Predicting the cookie quality of flours by using Mixolab®. *Eur. Food Res. Technol.* 227, 1549–1554. doi: 10.1007/s00217-008-0879-x
- Park, C. S., Kang, C. S., Cheong, Y. K., Jung, W., and Woo, S. H. (2010). Influence of puroindoline genotypes on grain characteristics, physico-chemical properties of flour and end-use quality of Korean wheats. *Breed. Sci.* 60, 233–242. doi: 10.1270/jsbbs.60.233
- Park, C.-S., Kim, H.-S., Kim, D.-H., Hyun, J.-N., and Kang, C.-S. (2012). Environmental impacts of Korean and CIMMYT wheat lines on protein characteristics and bread making quality. *Korean J. Crop Sci.* 57, 60–70. doi: 10.7740/kjcs.2012.57.1.060
- Patil, R. M., Tamhankar, S. A., Oak, M. D., Raut, A. L., Honrao, B. K., Rao, V. S., et al. (2013). Mapping of QTL for agronomic traits and kernel characters in durum wheat (*Triticum durum* Desf.). *Euphytica* 190, 117–129. doi: 10.1007/s10681-012-0785-y
- Payne, P. I., Holt, L. M., Jackson, E. A., Law, C. N., and Damania, A. B. (1984). Wheat storage proteins: their genetics, and their potential for manipulation by plant breeding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 304, 359–371. doi: 10.1098/rstb.1984.0031
- Payne, P. I., and Lawrence, G. J. (1983). Catalogue of alleles for the complex gene loci, Glu-A1, Glu-B1, and Glu-D1 which code for high-molecular-weight subunits of glutenin in hexaploid wheat. *Cereal Res. Commun.* 11, 29–35. doi: 10.1021/jf990151p
- Peña-Bautista, R. J. (2002). “Wheat for bread and other foods,” in *Bread Wheat: Improvement and Production*, eds B. C. Curtis, S. Rajaram, and H. G. Macpherson (Rome: FAO Plant Production and Protection Series).
- Preston, K. R., March, P. R., and Tipples, K. H. (1982). An assessment of the SDS-sedimentation test for the prediction of Canadian bread wheat quality. *Can. J. Plant Sci.* 62, 545–553. doi: 10.4141/cjps82-083
- Rasheed, A., Xia, X., Yan, Y., Appels, R., Mahmood, T., and He, Z. (2014). Wheat seed storage proteins: advances in molecular genetics, diversity and breeding applications. *J. Cereal Sci.* 60, 11–24. doi: 10.1016/j.jcs.2014.01.020
- Reif, J. C., Gowda, M., Maurer, H. P., Longin, C. F. H., Korzun, V., Ebmeyer, E., et al. (2011). Association mapping for quality traits in soft winter wheat. *Theor. Appl. Genet.* 122, 961–970. doi: 10.1007/s00122-010-1502-7
- Rubenthaler, G. L., Huang, M. L., and Pomeranz, Y. (1990). Steamed bread. I. Chinese steamed bread formulation and interactions. *Cereal Chem.* 67, 471–475.
- Shanhong, W., Fanhua, M., Li, Y., and Binghua, L. (2001). The effect of dwarf genes on agronomy characters in wheat. *J. Triticeae Crop.* 21, 5–9.
- Shewry, P. R., Halford, N. G., and Lafandra, D. (2003). Genetics of wheat gluten proteins. *Adv. Genet.* 49, 111–184. doi: 10.1016/S0065-2660(03)01003-4
- Slade, L., and Levine, H. (1994). Structure-function relationships of cookie and cracker ingredients. *Sci. Cookie Cracker Prod.* 9, 23–141.
- Tsilo, T. J., Simsek, S., Ohm, J. B., Harelend, G. A., Chao, S., and Anderson, J. A. (2011). Quantitative trait loci influencing endosperm texture, dough-mixing

- strength, and bread-making properties of the hard red spring wheat breeding lines. *Genome* 54:460470. doi: 10.1139/g11-012
- Van Der Borgh, A., Goesart, H., Veraverbeke, W. S., and Delcour, J. A. (2005). Fractionation of wheat and wheat flour into starch and gluten: overview of the main processes and the factors involved. *J. Cereal Sci.* 41, 221–237. doi: 10.1016/j.jcs.2004.09.008
- Veraverbeke, W. S., and Delcour, J. A. (2002). Wheat protein composition and properties of wheat glutenin in relation to breadmaking functionality. *Crit. Rev. Food Sci. Nutr.* 42, 179–208. doi: 10.1080/10408690290825510
- Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77
- Wang, F., Liu, T., Li, Q., An, Y., Xie, C., Sun, X., et al. (2017). QTL mapping of the pasting properties of wheat flour treated by papain digestion. *Starch Stärke* 69:1600077. doi: 10.1002/star.201600077
- Wang, L., Cui, F., Wang, J., Jun, L., Ding, A., Zhao, C., et al. (2012). Conditional QTL mapping of protein content in wheat with respect to grain yield and its components. *J. Genet.* 91, 303–312. doi: 10.1007/s12041-012-0190-2
- Würschum, T., Leiser, W. L., Kazman, E., and Longin, C. F. H. (2016). Genetic control of protein content and sedimentation volume in European winter wheat cultivars. *Theor. Appl. Genet.* 129, 1685–1696. doi: 10.1007/s00122-016-2732-0
- Yang, Y., Chai, Y., Zhang, X., Lu, S., Zhao, Z., Wei, D., et al. (2020). Multi-Locus GWAS of quality traits in bread wheat: mining more candidate genes and possible regulatory network. *Front. Plant Sci.* 11:1091. doi: 10.3389/fpls.2020.01091
- Yang, Y., Li, S., Zhang, K., Dong, Z., Li, Y., An, X., et al. (2014). Efficient isolation of ion beam-induced mutants for homoeologous loci in common wheat and comparison of the contributions of Glu-1 loci to gluten functionality. *Theor. Appl. Genet.* 127, 359–372. doi: 10.1007/s00122-013-2224-4
- Zhang, J., Dell, B., Biddulph, B., Drake-Brockman, F., Walker, E., Khan, N., et al. (2013). Wild-type alleles of Rht-B1 and Rht-D1 as independent determinants of thousand-grain weight and kernel number per spike in wheat. *Mol. Breed.* 32, 771–783. doi: 10.1007/s11032-013-9905-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Tian, Zhang, Li, Wen, Bi, Zhao, Wei, Chen, Yu, Shi, Liang, Xie, Li, Sun, Zhang and You. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genome-Wide Association and Genomic Prediction for Stripe Rust Resistance in Synthetic-Derived Wheats

Zahid Mahmood^{1,2}, Mohsin Ali³, Javed Iqbal Mirza⁴, Muhammad Fayyaz⁴, Khawar Majeed¹, Muhammad Kashif Naeem⁵, Abdul Aziz¹, Richard Trethowan⁶, Francis Chuks Ogbonnaya⁷, Jesse Poland⁸, Umar Masood Quraishi¹, Lee Thomas Hickey⁹, Awais Rasheed^{1,3*} and Zhonghu He^{3*}

¹ Department of Plant Sciences, Quaid-i-Azam University, Islamabad, Pakistan, ² Crop Sciences Institute, National Agricultural Research Centre (NARC), Islamabad, Pakistan, ³ Institute of Crop Sciences, CIMMYT-China office, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China, ⁴ Crop Disease Research Institute, NARC, Islamabad, Pakistan, ⁵ National Institute for Genomics and Advanced Biotechnology (NIGAB), National Agriculture Research Center (NARC), Islamabad, Pakistan, ⁶ Plant Breeding Institute, School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia, ⁷ Grains Research and Development Corporation, Kingston, ACT, Australia, ⁸ Department of Plant Pathology, Kansas State University, Manhattan, KS, United States, ⁹ Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Saint Lucia, QLD, Australia

OPEN ACCESS

Edited by:

Xue-Feng Ma,
Forage Genetics International,
United States

Reviewed by:

Frank Maulana,
Louisiana State University Agricultural
Center, United States
Shuyu Liu,
Texas A&M University System,
United States

*Correspondence:

Awais Rasheed
arasheed@qau.edu.pk
Zhonghu He
z.he@cgiar.org

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 02 October 2021

Accepted: 07 January 2022

Published: 24 February 2022

Citation:

Mahmood Z, Ali M, Mirza JI, Fayyaz M, Majeed K, Naeem MK, Aziz A, Trethowan R, Ogbonnaya FC, Poland J, Quraishi UM, Hickey LT, Rasheed A and He Z (2022) Genome-Wide Association and Genomic Prediction for Stripe Rust Resistance in Synthetic-Derived Wheats. *Front. Plant Sci.* 13:788593. doi: 10.3389/fpls.2022.788593

Stripe rust caused by *Puccinia striiformis* (*Pst*) is an economically important disease attacking wheat all over the world. Identifying and deploying new genes for *Pst* resistance is an economical and long-term strategy for controlling *Pst*. A genome-wide association study (GWAS) using single nucleotide polymorphisms (SNPs) and functional haplotypes were used to identify loci associated with stripe rust resistance in synthetic-derived (SYN-DER) wheats in four environments. In total, 92 quantitative trait nucleotides (QTNs) distributed over 65 different loci were associated with resistance to *Pst* at seedling and adult plant stages. Nine additional loci were discovered by the linkage disequilibrium-based haplotype-GWAS approach. The durable rust-resistant gene *Lr34/Yr18* provided resistance in all four environments, and against all the five *Pst* races used in this study. The analysis identified several SYN-DER accessions that carried major genes: either *Yr24/Yr26* or *Yr32*. New loci were also identified on chr2B, chr5B, and chr7D, and 14 QTNs and three haplotypes identified on the D-genome possibly carry new alleles of the known genes contributed by the *Ae. tauschii* founders. We also evaluated eleven different models for genomic prediction of *Pst* resistance, and a prediction accuracy up to 0.85 was achieved for an adult plant resistance, however, genomic prediction for seedling resistance remained very low. A meta-analysis based on a large number of existing GWAS would enhance the identification of new genes and loci for stripe rust resistance in wheat. The genetic framework elucidated here for stripe rust resistance in SYN-DER identified the novel loci for resistance to *Pst* assembled in adapted genetic backgrounds.

Keywords: GWAS, GBS, stripe rust (*Puccinia striiformis* Westend), synthetic hexaploid derived wheat, haplotype GWAS

INTRODUCTION

Stripe or yellow rust caused by an obligate pathogen *Puccinia striiformis tritici* (*Pst*) is a major threat to wheat production and grain quality. Wheat yield losses in different regions of the world up to 25% have been reported and this can climb to 80% when infections occur early in the crop season (Solh et al., 2012). Recently, stripe rust epidemics have damaged wheat production in many wheat growing countries and regions including Australia, Ethiopia, China, United States, Europe, South Africa, and South Asia (Milus et al., 2006; Chen, 2007; Wellings, 2011). Since Airborne *Pst* urediniospores can migrate to other regions of the world using the climatic system termed the “Western Disturbance,” thus, spreading new races. The Western Disturbance caused the spread of the (*Pst*) *Yr9* virulent race in the Indian Subcontinent and Nepal from the East African highlands between 1985 and 1997. In the past decade, virulence for *Yr27* caused epidemics in Pakistan and India on the commonly growing mega cultivars, Inqlab-91, and PBW-343, respectively (Duveiller et al., 2007).

Rust resistance, like other fungal diseases, can be controlled by fungicide and resistant cultivars. However, the use of a fungicide is associated with a high cost and is hazardous to the environment. Therefore, deploying resistant cultivars is environmentally friendly and particularly inexpensive for wheat growers. To date, more than 83 *Pst* resistance genes (*Yr1*–*Yr83*) have been catalogued in wheat and its wild relatives (Maccaferri et al., 2015; McIntosh et al., 2016). These are predominantly race-specific major genes, which interact with the pathogen according to the gene-for-gene model and produce hypersensitive reactions. This type of resistance is usually short lived when deployed in large areas; the evolution of new pathotypes of the pathogen population leads to a resistance breakdown. Virulence on *Yr2*, *Yr6*, *Yr7*, *Yr8*, *Yr9*, *Yr17*, and *Yr27* are examples of major gene resistance breakdown. It is essential that new sources of resistance are found and deployed to keep ahead of pathogen changes. However, minor genes or adult plant resistance (APR) genes are an alternative for major genes and provide a quantitative resistance that is often race non-specific and durable against various pathotypes.

Wheat breeders often rely on current or old varieties as a source of resistance, however, wheat wild relatives can also provide a useful source by direct recombination, bridge crosses, or including the development of synthetic wheats (Ogbonnaya et al., 2013). Within the wheat primary gene pool, considerable genetic variation exists in *Aegilops tauschii* and *T. turgidum* for resistance to both biotic and abiotic stresses (Halloran et al., 2008). The introgression of this genetic diversity through the development of synthetic hexaploid wheat (SHW) that can be directly crossed to adapted hexaploid wheat is one such strategy. Hexaploid wheat (SHWs) are known as primary synthetics and are generally obtained by artificially crossing of durum wheat (*T. turgidum*) and *Ae. tauschii*. These SHWs have been shown to carry genetic variation for resistance to numerous biotic and abiotic stresses (Mujeeb-Kazi et al., 1996; Ogbonnaya et al., 2013). The yellow, leaf, and stem rust resistance genes *Yr28*, *Lr21*, *Lr22*, *Lr32*, *Lr39*, *Lr41*, *Sr33*, *Sr45*, and *Sr46* were derived from *Ae.*

tauschii, and the *Sr* genes were subsequently shown to be resistant to the highly virulent *Ug99* race (Cox et al., 1995; Zegeye et al., 2014; McIntosh et al., 2016).

Genome-wide association studies (GWAS) are used to associate the genetic loci with phenotypic diversity (Huang and Han, 2014). This method combines a comparatively large portion of natural diversity in a species and localizes marker-trait associations to much shorter genomic regions because these diversity panels incorporate many more historical recombination events than classical recombinant inbred lines and doubled haploid populations (Nordborg and Weigel, 2008). The GWAS has proven to be a powerful tool for genetic analysis in wheat. It has been successful in identifying the genomic regions and markers for resistance to stripe rust in synthetic hexaploid wheat (Zegeye et al., 2014; Bhatta et al., 2019), global landraces collections (Jordan et al., 2015), Ethiopian durum wheats (Liu et al., 2017c), advanced lines derived from exotic crosses (Ledesma-Ramírez et al., 2019), Chinese wheat landraces (Long et al., 2019), global spring wheat collection (Maccaferri et al., 2015), global winter wheat collection (Bulli et al., 2016), US Pacific Northwest winter wheat (Naruoka et al., 2015; Liu et al., 2018), spring wheat (Muleta et al., 2017a), CIMMYT nurseries (Juliana et al., 2017), Afghan wheat landraces (Manickavelu et al., 2016), Ethiopian bead wheat (Muleta et al., 2017b), emmer wheat (Liu et al., 2017b), North American elite spring wheat (Godoy et al., 2017), elite ICARDA wheats (Jighly et al., 2015), diverse spring wheat (Kankwatsa et al., 2017), global landraces collection (Pasam et al., 2017), and elite durum wheat (Liu et al., 2017a).

Genome-wide prediction also referred to as genomic selection or genomic prediction is a technique to improve the selection accuracy and has the potential to reduce the cost of phenotyping and breeding cycles (Meuwissen et al., 2001) can help increase the rate of genetic gain especially in the case of quantitative traits. In the first step, genomic estimated breeding values (GEBVs) are estimated using a training set and different prediction models, and best prediction models are then used to select new germplasm developed by hybridization prior to field evaluation. The application of genomic prediction depends on the population size, marker density, model performance, heritability of the trait, training population size, and breeding population relatedness (Daetwyler et al., 2008; Bassi et al., 2016). In wheat, genomic prediction studies have been reported to predict rust resistance in diverse wheat landraces (Daetwyler et al., 2014; Crossa et al., 2016), landraces from Afghanistan (Tehseen et al., 2021), tetraploid wheat (Azizinia et al., 2020), and improved wheat germplasm (Ornella et al., 2012; Rutkoski et al., 2014; Bassi et al., 2016; Juliana et al., 2017).

This study was designed for: (i) evaluating the diversity for stripe rust resistance in 193 SYN-DERs against prevailing *Pst* races in Pakistan; (ii) conducting a GWAS analysis in SYN-DERs for resistance loci to the prevailing *Pst* races and identifying the linked SNP markers that could be deployed in marker-assisted selection (MAS); (iii) comparing genomic prediction accuracies for stripe rust resistance at seedling and adult plant stages using different models with two genotyping platforms, and (iv) determining whether some derivatives carry un-characterized genes for *Pst* resistance.

TABLE 1 | Virulence profile of Pst races used in this study.

Pathotype	Virulence on genes	Avirulence on genes
3.Pst.140202	Yr6, Yr7, Yr27, and YrExp2	Yr1, Yr5, Yr8, Yr9, Yr10, Yr15, Yr17, Yr24, Yr32, Yr43, Yr44, YrSp, YrTr1, and YrTye
5.Pst.173262	Yr6, Yr7, Yr8, Yr9, Yr15, Yr17, Yr27, Yr43, Yr44, and YrExp2	Yr1, Yr5, Yr10, Yr24, Yr32, YrSp, YrTr1, and YrTye
1.Pst.571242	Yr1, Yr6, Yr7, Yr8, Yr9, Yr17, Yr27, Yr43, and YrExp2	Yr5, Yr10, Yr15, Yr24, Yr32, Yr44, YrSp, YrTr1, and YrTye
4.Pst.571243	Yr1, Yr6, Yr7, Yr8, Yr9, Yr17, Yr43, YrExp2, and YrTye	Yr5, Yr10, Yr15, Yr24, Yr32, Yr44, YrSp, and YrTr1
2.Pst.571262	Yr1, Yr6, Yr7, Yr8, Yr9, Yr17, Yr27, Yr43, Yr44, and YrExp2	Yr5, Yr10, Yr15, Yr24, Yr32, YrSp, YrTr1, and YrTye

TABLE 2 | Mean response to *Puccinia striiformis* f. sp. tritici infection, estimates of variance components, and heritability.

Parameters	Islamabad (ISB)		Nowshera (NWS)		Across Locations	
	IT (0–9)	Severity (%)	IT (0–9)	Severity (%)	IT (0–9)	Severity (%)
Minimum	0.0	0.0	0.0	0.0	0.0	0.0
Mean	3.1	15.7	2.8	11.8	2.9	13.8
Maximum	8.5	90	8.5	80	8.5	85
σ^2_g	2.9***	278.6***	4.9***	247.7***	3.7***	255.8**
σ^2_e	0.9**	127.3*	1.9**	114.7**	1.5**	129.6**
σ^2_{ge}	3.8 ^{ns}	406.0 ^{ns}	6.8*	362.5 ^{ns}	5.3*	385.4 ^{ns}
σ^2_e	1.92	1.80	1.80	1.92	1.92	1.92
Heritability	0.75	0.68	0.71	0.68	0.72	0.66

σ^2_g , estimate of genotypic variance; σ^2_e , estimate of environmental variance; σ^2_{ge} , estimate of genotype $\times$ environment variance; σ^2_e , estimate of residual variance; H^2 , heritability; IT, infection type; DS, disease severity; ns, not significant; * $P < 0.05$; ** $P < 0.01$; and *** $P < 0.001$.

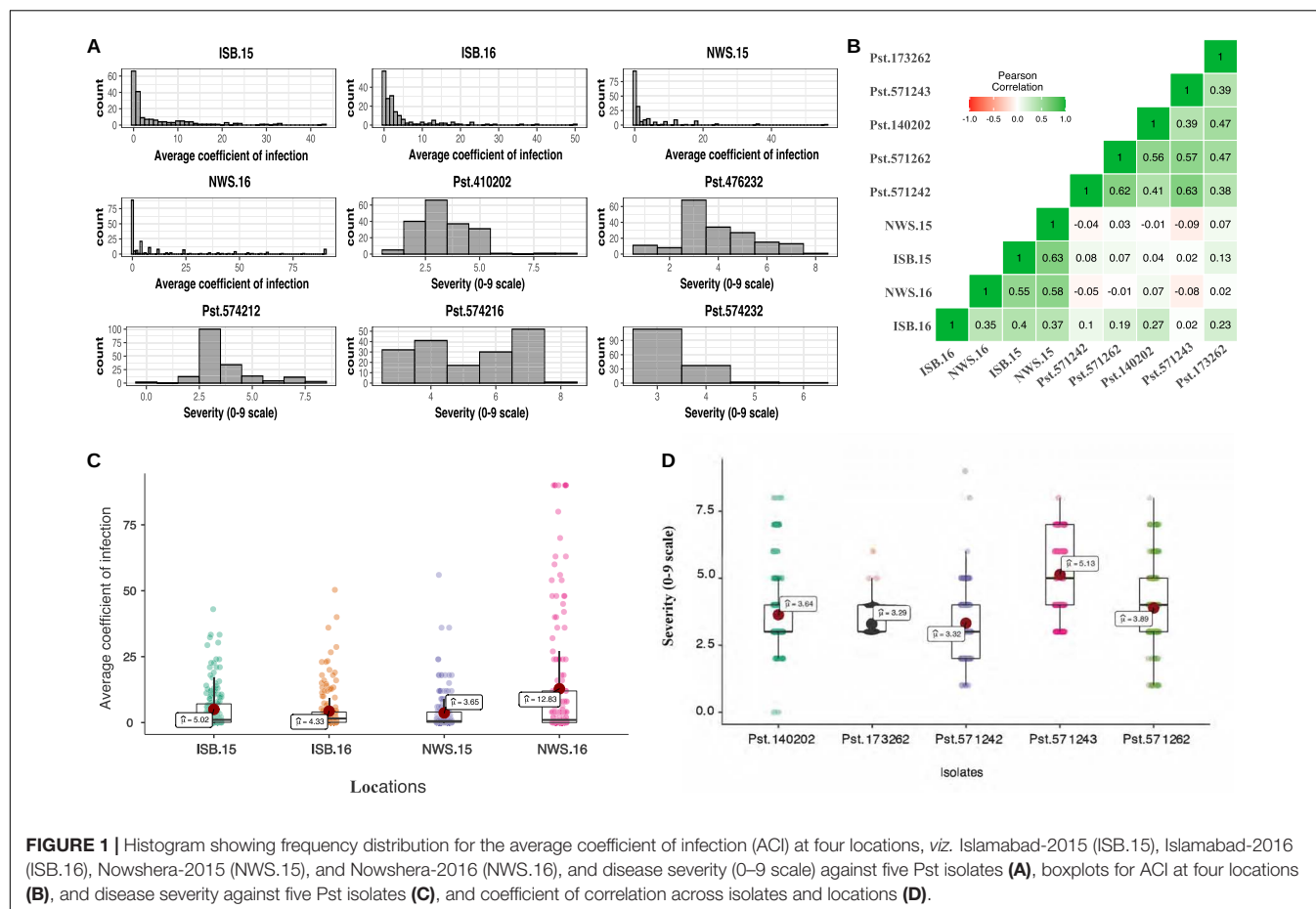


FIGURE 1 | Histogram showing frequency distribution for the average coefficient of infection (ACI) at four locations, viz. Islamabad-2015 (ISB.15), Islamabad-2016 (ISB.16), Nowshera-2015 (NWS.15), and Nowshera-2016 (NWS.16), and disease severity (0–9 scale) against five Pst isolates (A), boxplots for ACI at four locations (B), and disease severity against five Pst isolates (C), and coefficient of correlation across isolates and locations (D).

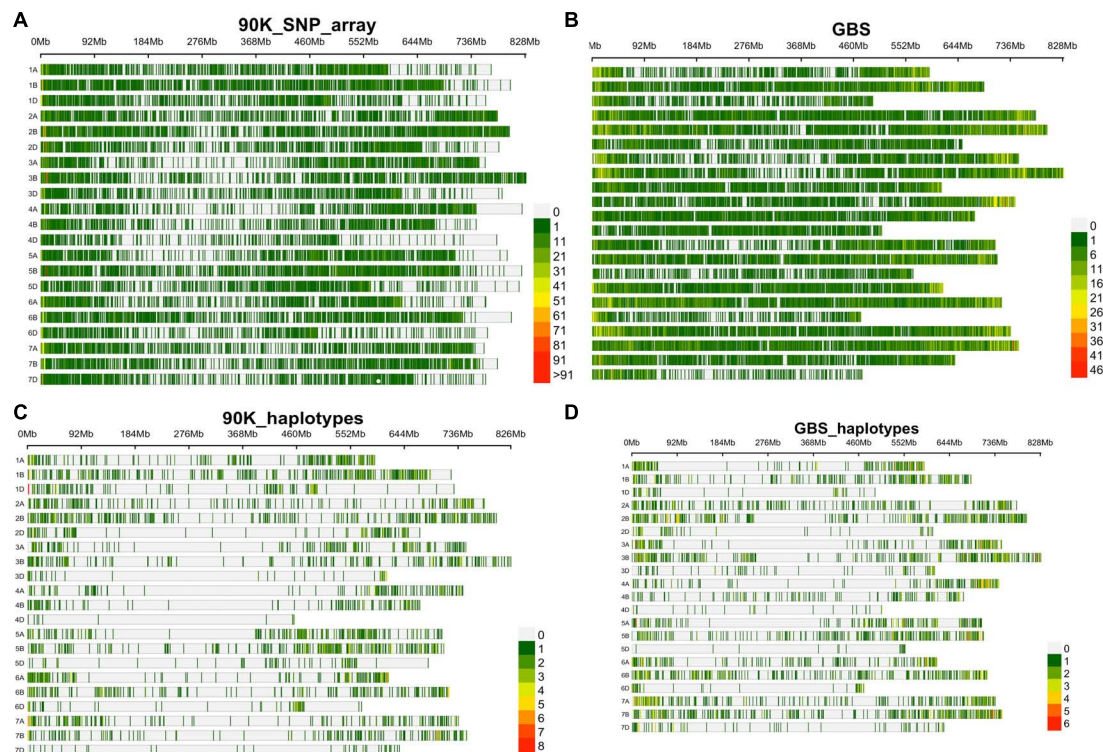


FIGURE 2 | SNP density and distribution in all 21 wheat chromosomes using 90K SNP array and GBS characterized in SYN-DERs, **(A)** in 90K SNP array, **(B)** in GBS, **(C)** haplotype density using 90K SNP array, and **(D)** haplotype density using GBS platform.

MATERIALS AND METHODS

Plant Materials and Experimental Sites

A panel containing 193 SYN-DERs were evaluated in this study (Supplementary Table 1). The details of the germplasm have been described earlier (Afzal et al., 2019). Briefly, the SYN-DERs were developed by crossing elite cultivars and advanced lines of spring wheat with synthetic hexaploidy wheats in several combinations (refer to a pedigree for details of primary synthetic hexaploid wheat accessions numbers). The field experiments were conducted at the National Agricultural Research Centre (NARC), Islamabad (33° 0'N, 73° 4'E) and Cereal Crop Research Institute (CCRI), Nowshera (34° 1'N, 72° 2'E) Khyber Pakhtunkhwa, Pakistan, in the winter field seasons of 2015–2016 and 2016–2017.

Seedling Stage Phenotyping

Seedling screening against stripe rust was performed at the Crop Disease Research Institute (CDRI), Murree, Pakistan under controlled conditions. Small plastic pots (8 cm × 10 cm) were filled with standard potting mix (soil and nursery substrate, 3:1), and were used to grow 5–6 plants of each accession including the susceptible wheat check cv. Morocco. The plants were grown in a glasshouse maintained at 50% humidity and 20°C. Genotypes were assessed for infection type responses to five *Pst* races: *Pst*.571242, *Pst*.571262, *Pst*.140202, *Pst*.571243, and *Pst*.173262

coded as Wang et al. (2016) and maintained at CDRI, Murree laboratory. These stripe rust races are frequently found in the yellow rust prone areas of Pakistan. The virulence and avirulence formulas for the isolates are provided in Table 1. The *Pst* isolates maintained at –80°C were heat shocked in a water bath at 42°C for 5 min. The mixture of petroleum ether (Merck Cat # 1.01775.2500) and paraffin oil (Merck Cat # 1.07162.1000) in a ratio of 4:1 was used to suspend the rust spores for inoculation on 10-day-old seedlings, at the two-leaf growth stage. The inoculum was applied using a fine mist atomizer. After inoculation, the mineral oil was allowed to evaporate, and the seedlings were then placed in a tray and watered. The *Pst* inoculated plant trays were shifted to a dark dew chamber at 100% relative humidity, 10°C temperature, and a light regime of 16 h light and 8 h dark for 24 h. Plants were then moved to a clean glasshouse under controlled temperature conditions of 15–18°C and 50% relative humidity. The same light/dark regime was continued during the rust evaluation. Water was non-limiting and recommended doses of liquid fertilizer were applied. Seedlings were treated with a growth inhibitor (Maleic Hydrazide) to slow plant development thus ensuring even disease infection and development. Notes on rust infection types were taken using a 0–9 scale (McNeal et al., 1971) on the 20th day of inoculation when susceptible genotype Morocco exhibited maximum infection. Seedling infection types (ITs) were classified as resistant with 0–4 (R), moderately resistant with score 5–6 (MR), and moderately to highly susceptible with score 7–9 (MS).

TABLE 3 | Haplotype blocks on wheat chromosomes, their number, block size, and number of SNPs per block using 90K SNP array and genotyping-by-sequencing (GBS) platform.

Chr	90K SNP array					GBS				
	N	Block size (kb)		SNPs		N	Block size (kb)		SNPs	
		Range	Mean	Range	Mean		Range	Mean	Range	Mean
1A	203	3–99871	10339	2–8	2.76	177	2–19011	99093	2–8	2.76
1B	302	8–98626	7957	2–11	3.03	232	2–17887	98552	2–7	2.68
1D	137	9–85350	7191	2–8	2.69	52	3–16674	98478	2–5	2.65
2A	205	8–96300	7101	2–9	2.79	238	2–15837	99436	2–8	2.6
2B	304	2–99707	8066	2–14	3.06	335	2–18030	99593	2–14	2.8
2D	120	9–99556	6590	2–8	2.72	54	18–16323	97876	2–5	2.41
3A	157	12–95781	9254	2–8	3.05	185	2–18778	99440	2–8	2.84
3B	204	4–95726	8039	2–12	3.14	342	2–19638	99474	2–11	2.76
3D	41	5–91287	11581	2–9	2.88	72	6–14033	97701	2–8	2.39
4A	136	11–78450	5730	2–14	2.84	181	2–20901	99366	2–8	2.91
4B	117	6–90285	7816	2–8	2.84	138	2–12206	98672	2–7	2.38
4D	10	13–91196	15167	2–5	2.8	23	3–11454	92585	2–6	2.43
5A	202	3–99690	8305	2–11	2.99	186	3–15977	97276	2–8	2.65
5B	283	8–95328	10391	2–13	3.2	271	2–20623	99815	2–7	2.82
5D	67	7–99753	10919	2–11	2.87	16	7–19401	91446	2–5	2.56
6A	179	4–92210	11380	2–10	3.02	190	2–14164	98516	2–11	2.65
6B	206	2–98128	6607	2–10	2.8	329	2–16501	99871	2–9	2.68
6D	62	10–93126	7480	2–6	2.68	46	3–15001	98599	2–6	2.52
7A	182	3–89043	8119	2–13	3.04	292	2–18888	99514	2–9	2.8
7B	169	6–95920	8702	2–19	3.24	364	2–17576	99805	2–10	2.64
7D	39	37–48011	5530	2–6	2.41	79	2–15770	99555	2–7	2.59

Adult Plant Disease Phenotyping

The diversity panel and a susceptible check (Morocco) were planted in 4 rows of 30 cm spacing and 2 m of length at NARC, Islamabad, and CCRI, Nowshera in 2015–2016 for screening for adult plant stripe rust resistance. The stripe rust susceptible cultivar Morocco was planted every 20th row to assist the spread of the rust epidemic. Inoculation was carried out using the *Pst* inoculum consisting of races used in this study. The inoculum was prepared by mixing rust spores mixture in liquefied petroleum ether (Merck Cat#1.01775.2500) and paraffin oil (Merck Cat # 1.07162.1000) in a ratio of 4:1 (V/V). The inoculum was sprayed with the help of a ULV sprayer on the rust spreader cultivar Morocco at the booting stage in both years at both field locations because this stage coincides with the favorable climatic conditions for rust spread. Rust infection and severity percentages were recorded when the genotype Morocco reached 70–80% severity. Rust scores were recorded three times each season at 1-week intervals to avoid disease escape. Wheat response to infection [infection types (IT)] was recorded using a 0–9 scale (Line and Qayoum, 1992). Yellow rust disease severity (DS) was noted as % infected leaf area of the host genotypes.

Analyses of Variance, Heritability, and Correlation

Analyses of variance of yellow rust infection types and disease severity from adult plant field evaluation were done across years

and environments using a linear mixed model to test for additive variance between genotypes, environments, and the interactions between genotypes by environments. In the mixed linear model, genotypes, and environments were used as fixed and years as random factors. Broad-sense heritability (H^2) was calculated using an ANOVA model to estimate variance components on a genotype mean basis.

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{y} + \frac{\sigma_e^2}{r}}$$

where, σ_g^2 is the genotypic variance, σ_e^2 is the environment variance, σ_{ge}^2 is the genotype by environment interaction variance, and σ_e^2 error is the residual error variance, y is the number of years, and r is the number of replications within each experimental site. Pearson correlation coefficients (r) among experimental sites and cropping seasons were estimated to examine the consistency of infection types and disease severity across the environments. Statistical analyses of the present study were performed using R Statistical Software.

DNA Extraction, and SNP Marker Genotyping

For genomic DNA extraction, five seeds of each SYN-DER accession were grown in 7 cm diameter disposable pots in a growth room. After 16–18 days of growth fresh leaf samples

TABLE 4 | Quantitative trait nucleotides (QTNs) associated with resistance to *Pst* races at seedling stage in SYN-DER panel using 90K and GBS markers.

Race	SNP	Alleles ^a	Chr	Pos	QTN effect ^b	LOD score	−log10(p)	r ² (%) ^c	MAF ^d
Pst.571243	IWB72742	<u>G</u> /A	1B	300.6	−0.74	4.58	5.36	15.39	0.21
Pst.571243	1B_338552631	T/ <u>G</u>	1B	338.6	1.05	3.74	4.48	9.18	0.06
Pst.571243	IWB73197	<u>T</u> /G	2B	152.2	−0.61	3.72	4.45	10.76	0.24
Pst.140202	2D_82307885	G/ <u>A</u>	2D	82.3	0.86	4.18	4.94	10.9	0.1
Pst.140202	3A_701489529	<u>C</u> /T	3A	701.5	−0.65	3.48	4.21	5.29	0.08
Pst.140202	3B_180646490	<u>T</u> /C	3B	180.6	−0.57	3.67	4.41	9.98	0.23
Pst.140202	IWB26725	<u>G</u> /A	3D	367.4	0.48	3.29	4.01	6.09	0.15
Pst.571262	IWB1577	T/ <u>C</u>	3D	439.7	−0.45	5.01	5.81	8.53	0.5
Pst.571242	IWB24288	<u>A</u> /G	3D	447.1	0.4	4.57	5.35	8.95	0.39
Pst.571243	4A_659618327	<u>T</u> /C	4A	659.6	0.55	3.61	4.35	11.3	0.42
Pst.571242	4B_11905357	G/ <u>A</u>	4B	11.9	0.48	6.88	7.74	11.93	0.3
Pst.571242	IWB5827	<u>T</u> /C	4B	603.1	−0.39	3.91	4.66	9.15	0.44
Pst.571243	4B_609362872	<u>A</u> /C	4B	609.4	−0.49	4.93	5.72	9.26	0.48
Pst.571242	5A_363980539	<u>A</u> / <u>G</u>	5A	364.0	−0.32	3.43	4.15	5.3	0.31
Pst.571242	5A_590355732	<u>C</u> /T	5A	590.4	−0.36	3.49	4.21	7.91	0.48
Pst.571242	IWB28556	<u>A</u> /G	5A	620.6	−0.37	3.61	4.35	7.13	0.31
Pst.140202	IWB27708	<u>A</u> /G	5B	2.3	−0.22	3.39	4.11	2.34	0.32
Pst.571242	IWA3089	C/ <u>T</u>	5B	580.4	0.43	3.29	4.01	10.6	0.47
Pst.571242	5B_580647907	<u>T</u> /C	5B	580.6	−0.75	7.24	8.11	10.52	0.08
Pst.571243	5B_580647907	<u>T</u> /C	5B	580.6	−0.77	4.19	4.96	7.19	0.09
Pst.571243	IWB35933	C/ <u>T</u>	5D	521.4	−0.62	3.4	4.12	8.94	0.22
Pst.173262	7A_529833812	<u>G</u> /C	7A	529.8	−0.2	3.06	3.76	7.6	0.16
Pst.571262	7A_696929784	G/ <u>T</u>	7A	696.9	−0.84	3.44	4.17	6.95	0.06

^aResistance allele is underlined.^bQTN effect is negative if minor allele is increasing phenotype and positive if major allele is increasing phenotype value.^cPhenotypic variation explained by the QTN.^dMinor allele frequency.

were taken to perform DNA extraction (Dreisigacker et al., 2013). Aliquant part of 50 µl DNA (50–100 ng/µl) for each sample was shipped in a 96-well plate arrangement for genotyping with high-density SNP markers, using the Wheat 90K SNP array (Wang et al., 2014), at the Department of Primary Industries, Victoria, Australia. The KASP marker for *Lr34/Yr18* was used to identify the durable rust resistance gene in SYN-DERs (Rasheed et al., 2016).

Genome-Wide Association Analyses Using SNPs and Haplotypes

The GWAS for stripe rust responses recorded in seedling and field experiments was performed by the multi-locus GWAS methods. The population structure was inferred from the principal component analysis (PCA), and PC scores from the first five principal components were used as a Q matrix. The kinship matrix (K) was calculated from the TASSEL version 5.0. Quantitative trait nucleotides (QTNs) were identified by meMLM (Wang et al., 2016) and FASTmrMLM (Tamba and Zhang, 2018) methods, which are included in the R-package mrMLM v 3.1.¹ For each trait, *P*-values were extracted from the TASSEL results. Manhattan and quantile-quantile (QQ; observed *P* values plotted against expected *P* values) were plotted using R package qqman (Turner, 2014).

¹<https://cran.r-project.org/web/packages/mrMLM/index.html>

The SNP linkage disequilibrium blocks (SNPLDBs) were constructed to identify the multiple alleles to fit the property of multiple alleles per locus in the SYN-DERs. The SNPLDB was constructed using RTM-GWAS software v1.2, which is publicly available at <https://github.com/njau-sri/rtm-gwas> (He et al., 2017). The output vcf from RTM-GWAS was used as a marker dataset for association analysis in TASSEL version 5.0.

Genomic Prediction Models

This study used 11 marker-based prediction models to assess predictability (i.e., the correlation between predicted and observed trait values) of unobserved phenotypes. All prediction models differed from each other with respect to assumptions regarding estimation of markers effects. Among the parametric models included, an MLM-based prediction model, a genomic best linear unbiased prediction (GBLUP), was computed using the “BGLR” function of R package Bayesian generalized linear regression (BGLR) version 1.0.8 (Pérez and De, 2014). The GBLUP prediction model utilizes a realized genomic relation matrix (G) to model correlation among individuals (Vanraden, 2008; Habier et al., 2013). In addition, the Bayesian linear prediction models, i.e., Bayesian Ridge Regression (BRR), Bayesian least absolute shrinkage, and selection operator (BL), BayesA, BayesB, BayesC, and reproducing kernel Hilbert spaces regression (RKHS) were also used in prediction analysis. These Bayesian prediction models were also used with the function

TABLE 5 | Quantitative trait nucleotides (QTNs) associated with resistance to *Pst* at adult plant stages in four environments in SYN-DER panel using 90K and GBS markers.

Environment	SNP	Alleles ^a	Chr	Pos (Mb)	QTN effect ^b	LOD score	−log10(p)	r ² (%) ^c	MAF ^d
ISB.15	IWB7628	T/C	1A	3.1	2.3	6.34	7.19	6.83	0.47
ISB.16	1A_3878168	G/T	1A	3.9	2.18	6.71	7.57	6.95	0.3
NWS.15	IWB4201	G/A	1A	4.0	−2.96	3.95	4.7	5.72	0.1
NWS.16	IWB21700	T/C	1A	534.3	−7.22	3.53	4.26	8.13	0.29
NWS.15	1A_560487941	G/A	1A	560.5	−2.35	4.52	5.29	5.98	0.17
NWS.15	IWB10188	G/A	1A	581.5	−3.08	5.11	5.91	9.09	0.15
ISB.15	1B_8591698	C/T	1B	8.6	−2.92	3.4	4.12	5.97	0.1
NWS.16	IWB64963	G/A	1B	86.8	7.81	5.69	6.51	11.1	0.48
ISB.15	IWB2120	A/C	1B	106.8	2.92	3.84	4.58	6.85	0.15
ISB.16	IWB49173	T/C	1B	327.8	3.11	5.19	6	10.57	0.19
ISB.15	1B_633336851	C/A	1B	633.3	2.65	3.74	4.48	3.26	0.09
ISB.15	1B_683306760	G/A	1B	683.3	−2.58	6.16	7	8.16	0.36
NWS.16	2A_566856454	G/T	2A	566.9	11.57	4.08	4.83	9.06	0.12
NWS.16	2B_163977776	G/A	2B	164.0	−9.38	5.72	6.55	8.33	0.18
NWS.16	2B_360129171	G/A	2B	360.1	−13.81	7.83	8.72	6.89	0.06
NWS.15	IWB35566	G/A	2B	783.2	−2.85	4.6	5.38	5.11	0.1
NWS.15	3A_130776756	C/T	3A	130.8	3.81	5.81	6.64	6.5	0.06
NWS.15	3A_503145562	A/G	3A	503.1	2.12	3.75	4.49	3.88	0.13
NWS.15	3A_736945971	A/T	3A	736.9	−2	5.46	6.28	7.22	0.41
ISB.15	IWA747	G/A	3B	55.5	−3.21	5.65	6.47	9.26	0.21
ISB.15	3B_55514953	T/C	3B	55.5	−6.08	14.26	15.27	29.31	0.18
NWS.16	3B_65339336	G/A	3B	65.3	−15.82	5.7	6.52	14.21	0.1
ISB.16	3B_470866042	A/G	3B	470.9	−2.83	4.72	5.51	4.71	0.09
ISB.16	3D_2620724	C/T	3D	2.6	−2.17	5.95	6.78	8.2	0.5
NWS.15	3D_355163225	T/C	3D	355.2	−1.94	3.36	4.08	3.69	0.16
ISB.16	3D_551073224	T/C	3D	551.1	−2.34	5.45	6.27	6.64	0.23
NWS.16	4A_438964494	C/T	4A	439.0	5.83	5.54	6.36	5.45	0.48
NWS.16	IWB68805	C/T	4A	733.6	4.91	3.35	4.07	3.75	0.3
ISB.16	4D_156687029	G/A	4D	156.7	−6.08	10.39	11.34	13.35	0.06
NWS.15	IWB33444	C/T	5A	481.9	−2.41	10.85	11.81	9.01	0.5
ISB.16	IWA4223	C/T	5A	670.4	−1.72	3.09	3.79	4.71	0.46
ISB.15	IWB7864	G/A	5B	2.6	−2.29	4.23	4.99	6.14	0.32
NWS.16	IWB65690	G/A	5B	10.8	8.23	6.46	7.31	12.44	0.49
NWS.15	IWB8592	G/A	5B	64.7	2.79	9.55	10.47	8.81	0.24
NWS.16	5B_207483057	G/A	5B	207.5	6.74	4	4.75	3.31	0.13
ISB.16	5B_471381890	A/G	5B	471.4	2.81	4.58	5.36	6.28	0.13
NWS.16	IWA2062	G/A	5B	542.6	−9.99	3.94	4.69	6.48	0.07
NWS.15	IWB65055	T/C	5B	692.6	−2.58	4.62	5.4	8.44	0.26
NWS.15	IWB14489	G/A	5D	133.5	−3.85	6.99	7.85	6.04	0.06
ISB.16	IWB9144	G/A	5D	487.6	−1.73	3.15	3.85	4.52	0.41
ISB.15	IWB30735	T/C	6A	297.7	−2.8	4.14	4.9	8.9	0.33
NWS.15	IWB66163	T/C	6A	415.9	−2.18	3.61	4.34	3.82	0.14
ISB.16	IWB40151	A/G	6A	546.6	−1.89	3.14	3.85	5.44	0.37
ISB.16	6A_595332866	T/C	6A	595.3	−2.64	3.94	4.69	4.1	0.09
ISB.16	IWB37028	T/C	6B	4.4	−5.97	11.61	12.58	14.01	0.06
NWS.16	6B_22858086	A/G	6B	22.9	−12.25	7.15	8.02	9.36	0.11
NWS.15	6B_31867138	C/T	6B	31.9	−4.97	9.13	10.05	10.2	0.06
ISB.15	6B_231490683	G/A	6B	231.5	2.46	4.29	5.06	5.98	0.25
NWS.16	6B_361469100	C/T	6B	361.5	−9.57	4.24	5.01	3.59	0.07
ISB.15	6B_419133836	G/A	6B	419.1	−3.17	6.3	7.14	6.62	0.15
NWS.15	6B_618067850	G/C	6B	618.1	−4.03	5.93	6.76	6.71	0.06

(Continued)

TABLE 5 | (Continued)

Environment	SNP	Alleles ^a	Chr	Pos (Mb)	QTN effect ^b	LOD score	−log10(p)	r ² (%) ^c	MAF ^d
NWS.15	6D_436810635	<u>G</u> /A	6D	436.8	−3.16	5.47	6.29	10.85	0.17
ISB.16	IWB74161	<u>C</u> /T	7A	47.0	−1.79	3.34	4.06	5.05	0.43
NWS.16	7A_234640959	C/ <u>T</u>	7A	234.6	5.9	4.71	5.49	5.49	0.43
ISB.16	IWB21762	<u>C</u> /T	7A	506.1	−2.21	4.53	5.3	6.44	0.28
ISB.16	7A_588284942	<u>G</u> /A	7A	588.3	1.78	3.86	4.61	4.34	0.3
NWS.16	7A_675526339	<u>G</u> /C	7A	675.5	−7.26	5.11	5.91	4.86	0.17
NWS.15	7A_676996750	<u>G</u> /A	7A	677.0	−3.29	5.28	6.08	7.02	0.09
ISB.16	IWB26214	<u>C</u> /T	7B	59.6	−3.19	6.87	7.73	8.24	0.14
NWS.15	IWA5939	T/ <u>C</u>	7B	582.3	4.04	12.21	13.19	11.9	0.15
ISB.15	IWB13912	<u>T</u> /C	7B	692.6	−3.28	5.45	6.26	7.47	0.12
NWS.16	IWB48256	<u>T</u> /C	7B	711.5	−6.84	3.59	4.32	7.87	0.35
ISB.16	IWB12163	<u>G</u> /A	7B	727.5	−2.06	5.62	6.44	6.45	0.4
ISB.16	7B_746448232	A/ <u>G</u>	7B	746.4	−1.87	3.49	4.22	3.87	0.2
NWS.16	IWB42068	<u>A</u> /G	7D	11.4	−5.84	3.69	4.43	6.05	0.4
ISB.15	IWB74163	<u>A</u> /C	7D	44.5	−3.07	9.65	10.58	12.08	0.41
ISB.15	IWB59266	<u>A</u> /G	7D	58.7	−3.11	4.95	5.74	7.59	0.14
ISB.16	7D_96173227	G/ <u>T</u>	7D	96.2	1.82	4	4.75	3.25	0.17

^aResistance allele is underlined.

^bQTN effect is negative if the minor allele is increasing phenotype and positive if the major allele is increasing phenotype value.

^cPhenotypic variation explained by the QTN.

^dMinor allele frequency.

“BGLR” in the R package “BGLR.” For the ridge regression BLUP (rrBLUP) model, we used the “mixed.solve” function from rrBLUP R package version 4.6 (Endelman, 2011). To implement the elastic net (EN), we used the “glmnet” function implemented in the glmnet R package version 2.0–18 (Friedman et al., 2010). To compute EN, the value of alpha.5 was used. The EN model is a combination of ridge regression and LASSO. In addition, non-parametric models, i.e., relevance vector machines (RVM) and Gaussian Processes (GP), were used to build a GS prediction model. The “rmv” and “gausspr” functions from kernlab R package version 0.9–27 (Karatzoglou et al., 2007) were used for RVM and GP modeling. To verify the predictability of the 11 models in the SYN-DER population, we evaluated the prediction accuracy by 10-fold cross-validation using a training set randomly apportioned into each fold. The data were partitioned into training population (90%) and validation population (10%) sets.

RESULTS

Phenotypic Variations for Stripe Rust Resistance in the SYN_DER Population

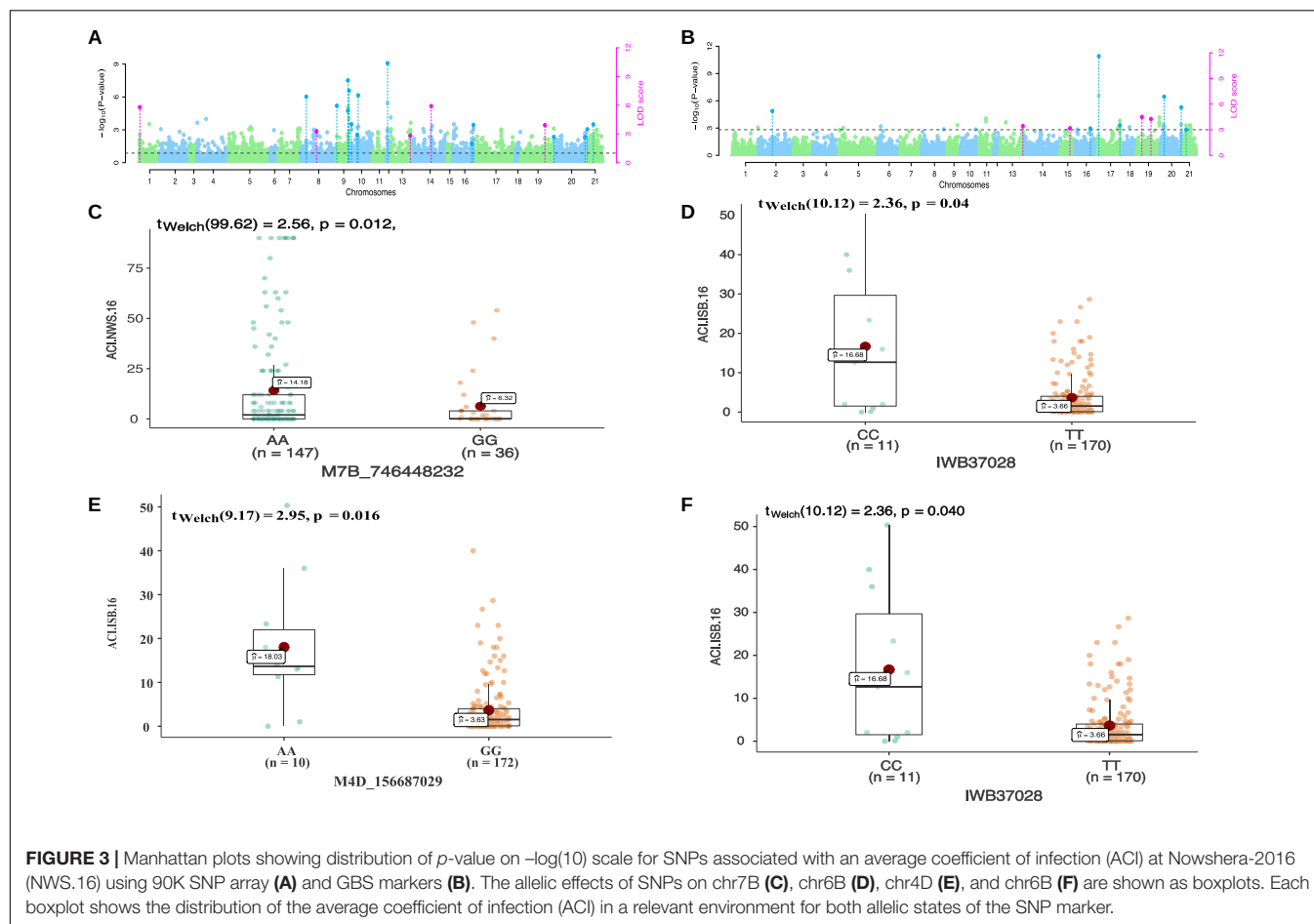
The response of the 193 SYN-DERs and check cultivars to *Pst* was assessed in four environments (two locations × 2 years) under high disease pressure. ANOVA showed highly significant ($P < 0.001$ and $P < 0.01$) differences among genotypes both for specific locations (Islamabad and Nowshera) and across locations (combined data) (Table 2). The variance components for environments were significant ($P < 0.01$ and $P < 0.05$). Similarly, genotypes by environment interactions for IT were

significant ($P < 0.05$) at Nowshera and across locations but non-significant for ISB.

The frequency distribution of ITs displayed by the SYN_DERs in response to the five *Pst* races is presented in Figures 1A,B. Of the lines tested, 78% (152), 63% (122), 79% (153), 38% (75), and 80% (156) of accessions showed seedling resistance to *Pst.571242*, *Pst.571262*, *Pst.140202*, *Pst.571243*, and *Pst.173262*, respectively. Notably, 18 (9%) of the genotypes showed seedling resistant infection types to all five *Pst* races.

The population showed a wide range of ITs across the environments. At Nowshera, 12% of accessions (24 genotypes) were highly resistant, 21% (41 genotypes) were showed resistant reactions, and 3 genotypes (1.5%) were highly susceptible. At ISB, 3% (6 genotypes) were highly resistant, 12% (24 genotypes) were resistant, and 1% (2 accessions) showed highly susceptible reactions (Figure 1). Eighteen (9%) of accessions were resistant in both cropping seasons at both experimental locations. Broad sense heritability (H^2) for IT and disease severity ranged from 0.66 to 0.75% (Table 2).

Pearson correlation coefficients between stripe rust IT and disease severity between Islamabad and Nowshera in both years are presented in Figure 1E. Correlations were 0.51 and 0.59 for ITs, and 0.38 and 0.61 for disease severity at Islamabad and Nowshera, respectively. The correlations between Islamabad and Nowshera for ITs in 2015 and 2016 were 0.65 and 0.40, respectively. The respective disease severity correlations were 0.63 and 0.30. All five *Pst* races evaluated for ITs were significantly and positively correlated to each other and values ranged from 0.34 to 0.61 (Figure 1D). Seedling infection types and disease severity of *Pst.140202* and *Pst.173262* were positively and were significantly correlated with adult plant ITs and disease severity in Islamabad in 2016.



SNP and Haplotype Variations in the Synthetic-Derived Diversity Panel

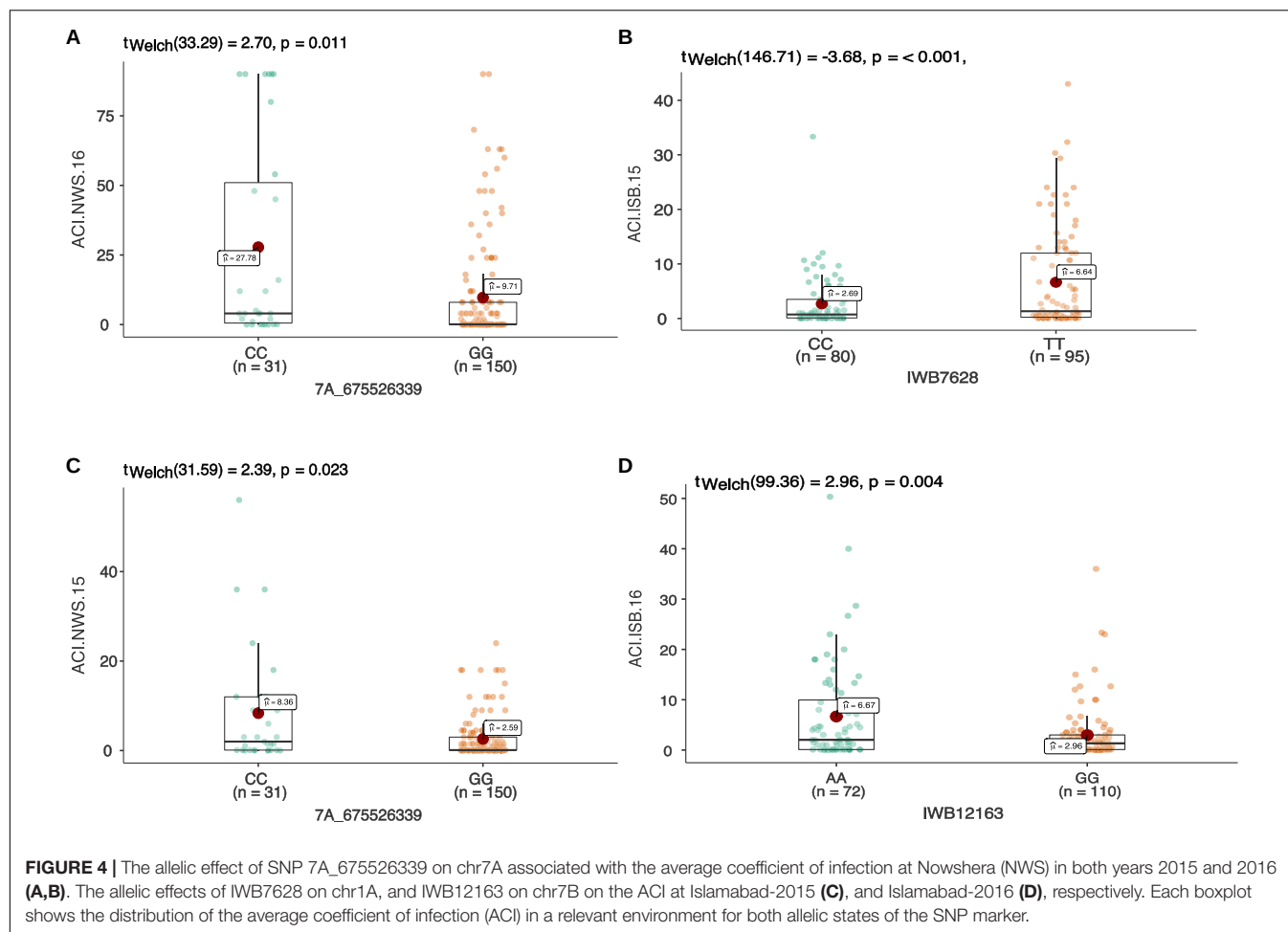
Two genotyping platforms, 90K SNP array and GBS, were used for GWAS. For the 90K SNP array, 29,632 SNP markers were retained after removing SNPs with missing data of >10% and minor allele frequency of <0.05. **Figure 2A** shows SNP marker density on each wheat chromosome. For GBS, out of 236,327 SNPs identified, 47,122 were finally used after removing SNPs with >10% missing data, and <5% minor allele frequency (**Figure 2B**).

Haplotype blocks were constructed using both genotyping platforms using the block partitioning approach with CIs based on genome-wide LD (D') patterns (Gabriel et al., 2002), and implemented in RTM-GWAS (He et al., 2017). **Table 3** describes the number of haplotype blocks, the range and average size of blocks in terms of kb, and the range and average number of SNPs comprising each haplotype block on each chromosome. In the 90K SNP array, 19,070 LD blocks were constructed (**Figure 2C**), out of which 3,325 blocks contained more than two haplotype (alleles) (**Table 3**). The maximum number of haplotypes ($n = 304$) were constructed on chr2B, while the minimum was on chr4D ($n = 10$). On an average, the haplotype block size ranged from 5.5 Mb (chr7D) to 11.5 Mb (chr3D). The number of SNPs in each

haplotype block was minimum 2 and maximum 14. In GBS, the number of blocks ranged from 16 (chr5D) to 364 (chr7B) (**Figure 2D**). The haplotype block size ranged from 2 to 19.6 Mb (chr3B), while SNPs/block ranged from 2 to 11 (chr3B and chr6A).

Association Analysis for Seedling Resistance to *Puccinia striiformis* in Synthetic-Deriveds

In total, 23 QTNs were identified for seedling resistance against five races in the SYN-DERs populations (**Table 4**). Eight QTNs were associated with seedling resistance against Pst.571242, of these the QTN on chr5B at 580.6 Mb was identified by both 90K SNP array and GBS and accounted for 10.5% of the total phenotypic variation. Only two QTNs were identified for resistance against Pst.571262 on chr3D and 7A, and explained 8.5 and 6.9% of the total variation, respectively. Five QTNs were detected against Pst.140202: these explained 2.3 to 10.9% of the total variation and were distributed on chr2D, chr3A, chr3B, chr3D, and chr5B. Seven QTNs were identified for resistance against Pst.571243 and accounted for 7.9 to 15.3% of the total variation. Only one QTN was identified for resistance against Pst.173262 on chr7A and explained 7.6% of the total variation.



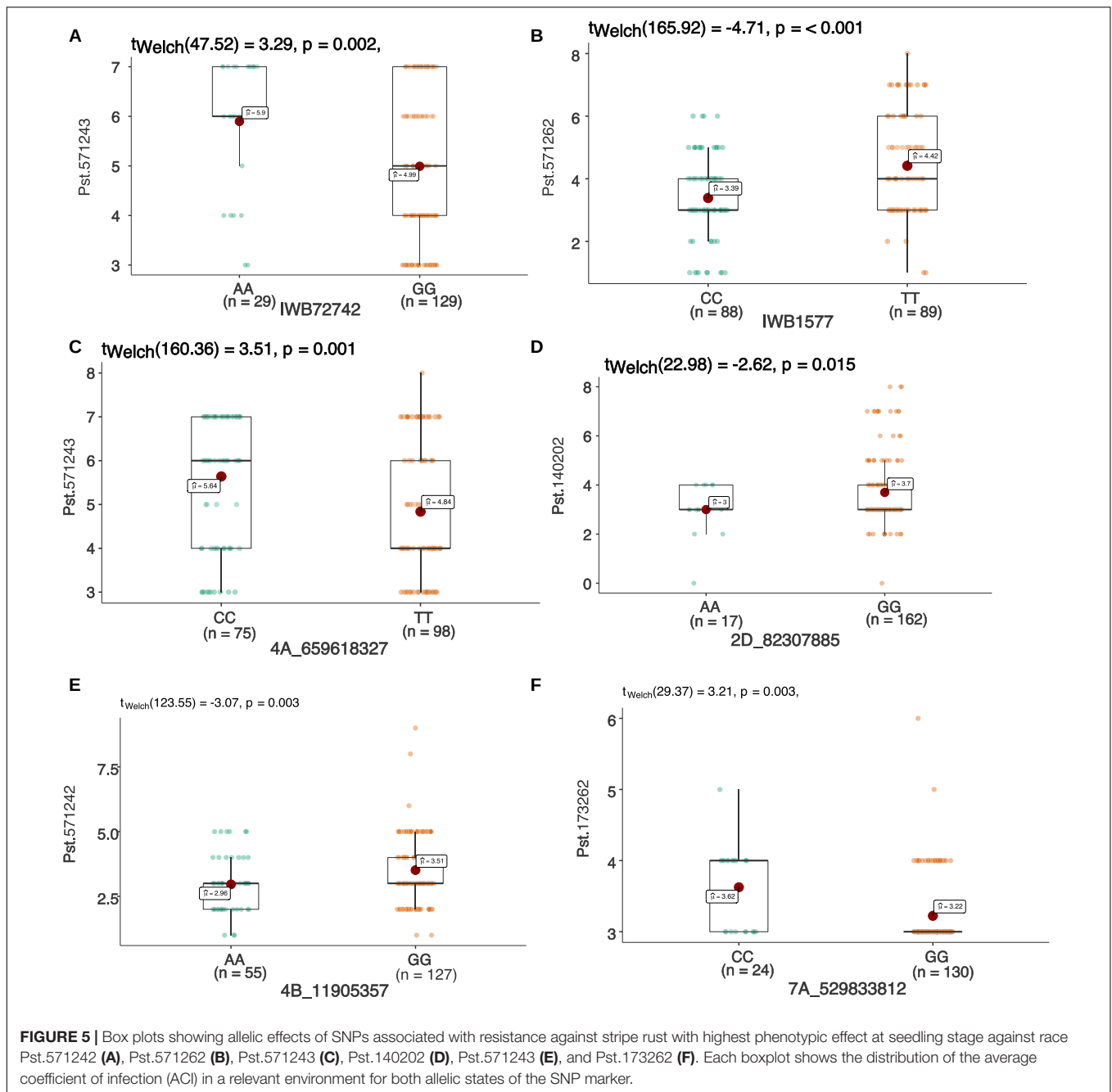
Association Analysis for Adult Plant Resistance to *Puccinia striiformis*

In total 68 QTNs were identified for adult plant resistance against Pst in SYN-DERs (Table 5). Fourteen QTNs were identified for ISB.15, 19 for ISB.16, 18 for NWS.15, and 17 for NWS.2016. These QTNs were detected on all chromosomes except chr1D, chr2D, and chr4B. Figure 3 shows Manhattan plots for significant SNPs associated with resistance to Pst at NWS.16 using a 90K SNP array (Figure 3A) and GBS (Figure 3B). The allelic effects of associated SNPs are shown as box plots (Figures 3C–F). The phenotypic variation explained by the QTNs ranged from 3.2% (96.1 Mb at chr7D) to 29.3% (55.5 Mb at chr3B). Some QTNs were identified by both genotyping platforms, i.e., at 55.5 Mb on chr3B for resistance to Pst at ISB.15, and 560.4 Mb at chr1A for NWS.15. A QTN at 53.4–58.1 Mb on chr1A associated with NWS.15 and NWS.16, and another on chr7A at 675–676 Mb associated with NWS.15 and 16 (Figures 4A,B). Similarly, QTN on chr7B at 711–727 Mb was associated with resistance to Pst at NWS.16 and ISB.16 (Figure 4C). Some QTNs were associated with resistance to Pst at multiple environments including QTN at 3.1–3.9 Mb on chr1A associated with ISB.15, ISB.16, and NWS.15 (Figures 4C,D). Interestingly some QTNs were associated with both seedling and adult plant resistance,

i.e., the QTN on chr1B at 300–327 Mb was associated with Pst.571243 at ISB.16 (Figure 5A), and the QTN at 152–163 Mb on chr2B associated with Pst.571243 at NWS.16. Similarly, a QTN at 355–367 Mb on chr3D was associated with Pst.140202 at NWS.15 (Figure 5B). The QTNs of chr4A, chr2D, and chr4B were associated with Pst.571243, Pst.140202, and Pst.571242, respectively (Figures 5C–E). A QTN on chr7A at 675–696 Mb identified in NWS.15 and 16 was also associated with Pst.571262 (Figure 5F). The allelic effects were also determined for the durable rust resistance gene Yr18, and the resistance allele was significantly associated with resistance to Pst in all four environments, i.e., ISB.16 (Figure 6A), NWS.16 (Figure 6B), ISB.15 (Figure 6C), and NWS.15 (Figure 6D).

Haplotype Blocks Associated With Resistance to *Puccinia striiformis* at Seedling and Adult Plant Stages

In total, three haplotype blocks were associated with seedling resistance against Pst.571242, Pst.140202, and Pst.173262 on chr1A, chr3B, and 7D, respectively (Table 6). The haplotype block on chr1A identified with the 90K SNP array was present at 575.2 Mb and contained six haplotypes, whose frequency ranged from 1.03 to 59% (Figures 7A,B). The effect of all three



haplotypes of chr1A LD block is shown in **Figure 7C**. A haplotype block on chr3B by (GBS markers) was positioned at 125.8 Mb and had four haplotypes with a frequency of 67 to 1%. Similarly, the haplotype block on chr7D was present at 627.3Mb and contained five variants with a frequency between 1.5 to 79.2%. This haploblock is likely a homolog of the QTN identified on chr7A for resistance against the same race.

In total, 11 haplotype blocks (two identified with GBS and nine with 90K SNP array) were associated with *Pst* resistance at the adult plant stage. Two haplotype blocks on chr5A and chr1A were associated with resistance to *Pst* at ISB.16, with

three and six haplotype variants observed, respectively. The haplotype block on chr6B was associated with resistance to *Pst* at NWS.16 and consisted of three haplotypes (**Figures 8A,B**), where the Hap-II (CCG) significantly reduce the ACI (**Figure 8C**). Similarly, haplotype block on chr1A consisted of four haplotypes (**Figures 8D,E**), and Hap-II (CATCTCTCA) was associated with resistance to *Pst* at NWS.15 (**Figure 8F**). A haplotype block at 488 Mb on chr1D was associated with resistance to *Pst* at ISB.15 and NWS.15, while another haplotype block on chr5A at 465 Mb was associated with *Pst* resistance at ISB.16 and NWS.15. This haplotype block is likely the QTN at 481 Mb which was associated

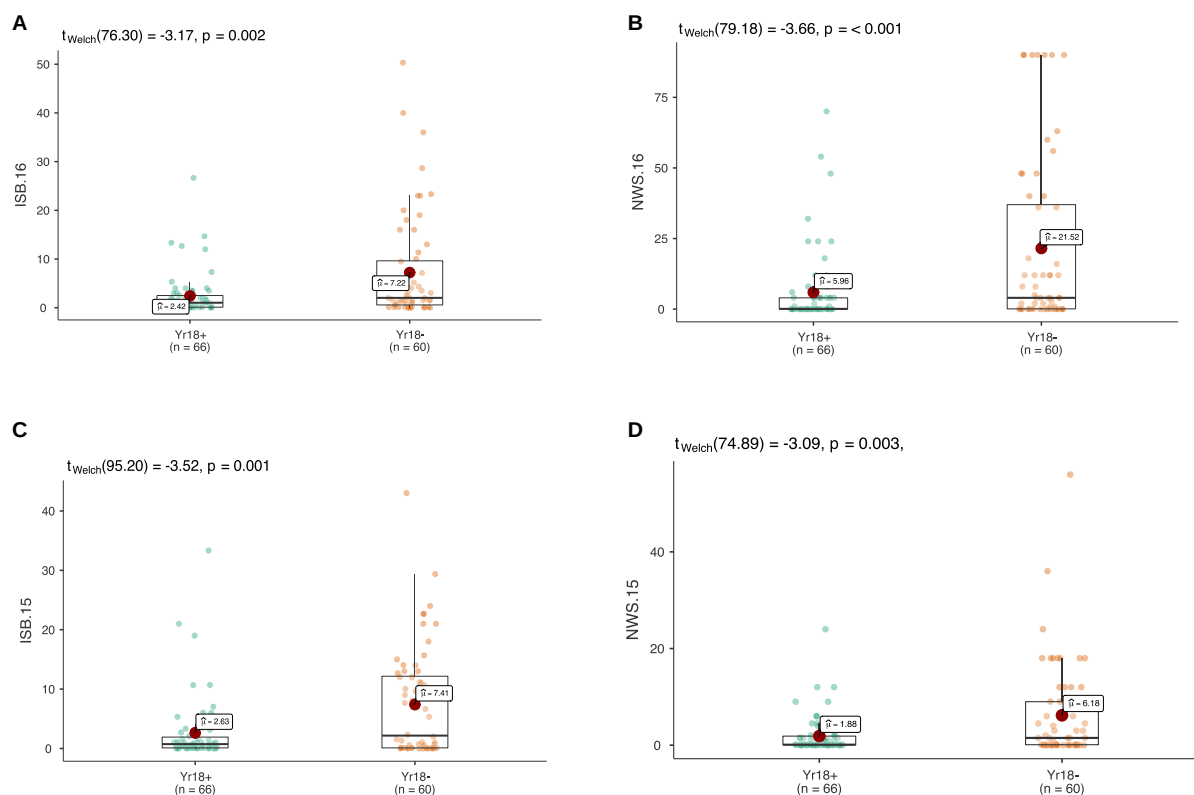


FIGURE 6 | The allelic effects of the durable rust resistance gene *Lr34/Yr18* on the average coefficient of infection (ACI) in four environments at adult plant stage in ISB.15 (A), ISB.16 (B), NWS.15 (C), and NWS.16 (D). Each boxplot shows the distribution of the average coefficient of infection (ACI) in a relevant environment for both allelic states of the SNP marker.

with APR at NWS.15. Five haplotype blocks were associated with *Pst* resistance at NWS.15 and these were distributed across chr1D (2), chr2B, chr5A, and chr6A. For the haplotype block on chr2B (6.2 Mb), 13 different haplotype variants were identified with a frequency ranging between 1.03 and 43%.

Genomic Prediction for Resistance Against *Puccinia striiformis*

Genomic-prediction analysis was conducted using a fivefold validation for *Pst* resistance at four locations and five *Pst* races using 11 different prediction models (Table 7). In the case of APR, prediction accuracies ranged from 0.23 (RKHS for ISB.15) to 0.511 (BL for NWS.15) using 90K markers, while prediction accuracies were relatively lower for GBS. Among the prediction models, BRR, BL, and GBLUP showed higher prediction accuracies compared to other models. Prediction accuracies were low for *Pst*.173262 and *Pst*.140202 using both GBS and 90K markers. The hierarchical clustering was used to classify the prediction models, which indicated that EN-based prediction accuracies were quite different than other models both for 90K and GBS markers (Figures 9A,B). Based on the 90K platform, all Bayes model (A, B, and C) and BL were quite similar

in the prediction of reaction against *Pst*. BRR, GP, and GBLUP were quite similar in the case of the GBS platform (Figure 9A), while BRR was a bit different compared to GP and GBLUP in the case of 90K markers (Figure 9B).

DISCUSSION

Stripe Rust Resistance in Synthetic-Deriveds at Seedling and Adult Plant Stages

The deployment of new, effective, and durable sources of resistance against *Pst* is required to reduce the risk of epidemics. Seven SYN-DERs were found to possess a high level of resistance against three *Pst* races, while six were resistant against all five races. It is likely that these SYN-DERs (SD37, SD38, SD73, SD85, SD104, SD172, and SD173) carry major stripe rust resistance genes. All five races used in the evaluation were avirulent to *Yr24/Yr26*, which was identified in synthetic hexaploid wheats and has been deployed in China and elsewhere (McIntosh et al., 2018). Most *Pst* races are avirulent to the *Yr24/Yr26* gene, however, races virulent to *Yr10* were also virulent to *Yr24/Yr26*,

e.g., Australian *Pst* race 150 E16A + and Chinese *Pst* races V26-CH42, and V26-Gui22 (McIntosh et al., 2018). Since SYN-DER wheats are not extensively deployed in Pakistan and the races used in this study were the most virulent available races, it is likely that virulence to *Yr24/Yr26* is not common in the pathogen population in Pakistan. Therefore, the eight SYN-DERs could be an excellent source of resistance against *Pst* in Pakistan and other countries where virulence to this gene combination is not present.

At the adult plant stage, more than 110 SYN-DERs showed moderate to resistant responses against *Pst*. The field screening was carried out in 'hot-spot' areas of *Pst* incidence, thus, this APR in SYN-DER could be usefully deployed against *Pst* races in the region. These results are in accordance with previous findings that APR occurs at a high frequency in synthetic hexaploid wheat (Zegeye et al., 2014; Bhatta et al., 2019). This is partly attributable to the fact that the A and B genomes of durum wheat are present completely in synthetic hexaploid wheat and partially in SYN-DERs. Previous studies indicate that *Pst* isolates from bread wheat are often avirulent on durum wheat (Aoun et al., 2021). Among the APR SYN-DERs, 68 carried the *Lr34/Yr18* gene, which is known to provide a partial resistance against all *Pst* races. The results also suggested the presence of *Lr34/Yr18* reduced overall incidence of *Pst* in all four environments against all five isolates. However, none of the SYN-DERs carried *Lr67/Yr46*, which was expected because this gene evolved after polyploidization and is mostly present in landraces from Pakistan and India (Riaz et al., 2016), while synthetic hexaploid wheats and parents used in SYN-DER did not have any introgression from Pakistan or Indian landraces.

Quantitative Trait Nucleotides and Haplotypes Associated With *Puccinia striiformis* Resistance in Synthetic-Deriveds

Both platforms, i.e., GBS and 90K SNP array, effectively identified the loci associated with resistance to *Pst*, and some QTNs were common to both platforms. We have collected information for stripe rust resistance loci from 35 different studies (Supplementary Table 2) and compared our QTNs with previous findings. Among the *Yr* resistance genes, *Yr24/Yr26* is derived from synthetic wheats and widely deployed in synthetic wheat-based commercial cultivars in China (Zeng et al., 2014). Previously, GWAS identified several *Yr* resistance loci co-localized with known *Yr* genes including *Yr24/Yr26/Yr28* on chromosome 1B, *Yr48* on chromosome 5AL, *Yr32* on chromosome 2A, and *Yr19* on chromosome 5BL (Zegeye et al., 2014). Apart from *Yr24/Yr26*, it was expected that several of the SYN-DERs could carry *Yr32* because all five races are avirulent to this gene. One QTN and one haplotype were associated with resistance to *Pst* on chr2A at 566 and 30.8 Mb, respectively. The QTN at ~566 Mb was likely to be *Yr32*; previously, the SNP AX-108752496 (similar position) was reported to be associated with *Pst* resistance (Wu et al., 2021). However, the minor allele provided resistance and its frequency at this QTN was 12% ($n = 23$), fourteen out of 23 SYN-DERs also possessed the durable rust resistance gene

Lr34/Yr18. Therefore, these 14 accessions could carry both major and minor genes, thus, provide valuable donor sources for breeding programs.

Among the 32 seedlings and 68 APR QTNs, 18 had a phenotypic effect exceeding 10%. The largest effect QTN on chr3B at ~55.4 Mb explained 29.3% of the variation was identified by both platforms in two environments. However, the phenotypic variation explained was relatively lower in ISB.16 (14.2%). Yao et al. (2020) previously identified a QTN at a similar position in Chinese wheat landraces (designated *QYr.nafu.3BS*). Since the major allele provided resistance at this QTN, it is likely that this locus was responsible for the high frequency of the resistant SYN-DERs.

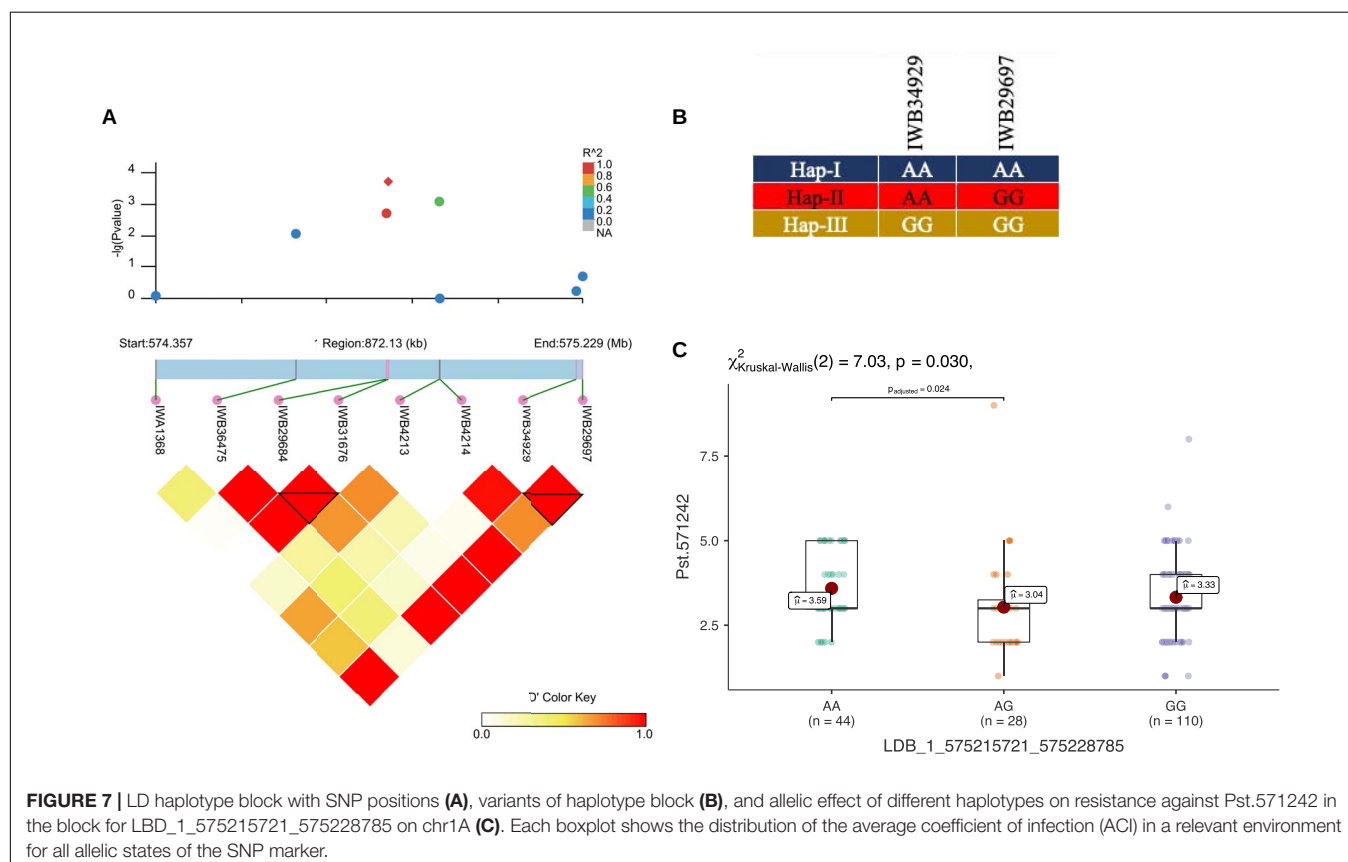
Previous studies identified QTNs for resistance to *Pst* on chromosomes 2A, 3B, 6A, and 7B in an association mapping panel of 181 SHWs (Zegeye et al., 2014). The QTNs and haplotypes identified on the D genome showed the potential of SYN-DERs for improving the *Pst* resistance in modern wheat cultivars. The same loci associated with seedling and APR to *Pst* on chr1B (~300 Mb), chr7A (~506 Mb), chr2B (~150 Mb), chr3A (~701 Mb), chr3D (~355 Mb), and chr5B (~2.2 Mb). These loci could be used to discover potentially novel alleles of major stripe rust resistance genes. The genes *Yr18*, *Yr29*, *Yr30*, and *Yr78* have been widely used in wheat breeding (Wu et al., 2021). However, in our study, no SNP association was found in the vicinity of *Yr78* and *Yr30*. The QTN on chr1B at 683 Mb was likely to be *Yr29*, and a QTL *QYr.nwafu-1BL* was also identified in close proximity (Wu et al., 2021). The QTN on chr3A was identified as effective against *Pst*.140202, and a major gene *Yr75* is located nearby at ~675 Mb, while the stem rust resistance gene *Sr15* was identified at the same position (Babiker et al., 2015). However, several loci identified in this study could not be compared with the previous studies due to the absence of a meta-analysis of stripe rust resistance loci in wheat. The establishment of such a framework would greatly enhance the validation and identification of loci associated with stripe rust resistance, particularly in GWAS studies.

Our study applied an LD based haplotype approach to discover loci associated with resistance to *Pst*. Until now, only one haplotype-based GWAS for stripe rust resistance has been reported (Wu et al., 2021). Previously, we used a haplotype-GWAS approach in SYN-DERs to identify the loci associated with drought adaptability (Afzal et al., 2019). The results confirmed that haplotype-GWAS was an effective strategy to increase the power of GWAS experiments. Here, we showed that haplotype-GWAS identified 9 out of 13 trait-associated loci where individual SNPs were ineffective. This was because haplotypes containing a group of closely linked SNP markers can increase the level of polymorphisms and overcome the limitation of using single SNP markers by creating more combinations (haplotypes). Several haplotypes associated with phenotypes in our study were not identified by SNP-GWAS and this could be due to many factors, including patterns of LD in the population, marker density, and the genetic architecture of the trait. The haplotype on chr6B (Figures 7A–C) associated with APR was also identified by SNP-GWAS and is likely to be a new locus. Similarly, a haplotype

TABLE 6 | Haplotypes associated with resistance to *Pst* at seedling and adult plant stages in SYN-DER wheats using 90K and GBS markers.

Trait	Haplotype ID	Chr	Position	SNPs/ block	-log ₁₀ (p)	Genotype (Frequency)		
						Hap-I	Hap-II	Hap-III
ISB.16	LDB_1_25490035_25490120	1A	25490035	2	1.45E-06	GA (0.68)	AG (0.30)	GG (0.019)
Pst.571242	LDB_1_575215721_575228785	1A	575215721	2	1.00E-08	GG (0.59)	AA (0.18)	AG (0.15)
NWS.15	LDB_3_19112718_19129042	1D	19112718	2	1.19E-06	AA (0.74)	TG (0.25)	
ISB.15	LDB_3_488576303_488577792	1D	488576303	4	4.00E-07	ATGT (0.55)	GGAC:0.37	GGGT (0.02)
NWS.15	LDB_3_488576303_488577792	1D	488576303	4	1.00E-08	ATGT (0.55)	GGAC:0.37	GGGT (0.02)
ISB.15	LDB_4_30830742_30831056	2A	30830742	2	7.00E-08	CA (0.92)	TG:0.06	CG (0.01)
NWS.15	LDB_5_6258683_6338084	2B	6258683	9	3.00E-12	CATTCTTCA (0.43)	CACCCTTCA (0.18)	TGTTCTCG (0.14)
Pst.140202	LDB_8_125880000_125930410	3B	125880000	4	6.91E-06	GACT (0.67)	GGTC (0.15)	AGTC (0.15)
NWS.16	LDB_10_111292188_111292941	4A	111292188	2	2.00E-09	GC (0.47)	AT (0.44)	GT (0.046)
NWS.15	LDB_13_465541110_465541233	5A	465541110	2	5.00E-07	CG (0.58)	TA (0.35)	TG (0.02)
ISB.16	LDB_13_465541110_465541233	5A	465541110	2	1.00E-08	CG (0.58)	TA (0.35)	TG (0.02)
NWS.15	LDB_16_584678556_584680439	6A	584678556	2	8.00E-06	GT (0.60)	TG (0.31)	GG (0.025)
NWS.16	LDB_17_15781175_15781777	6B	15781175	3	1.00E-07	CCG (0.34)	CTG (0.31)	TCT (0.25)
Pst.173262	LDB_21_627325333_627325482	7D	627325333	2	6.00E-07	GC (0.79)	AT (0.103)	GT (0.06)

*Bold haplotype blocks are the loci also identified by SNP-GWAS. Only the top three most frequent haplotypes in each LD block are mentioned, and the values in parentheses are the frequencies of the relevant haplotypes in the diversity panel.



consisting of nine SNPs on chr2B at 6.2 Mb (Figures 7D–F) was not identified by SNP-GWAS, and Hap-II, which provided a high level of APR was present in 83 accessions. Several genes and QTL have been identified on chr2B including *Yr32*, *Yr43*, *Yr44*, *Yr53*, *Yr72*, *Qyr.cim.2BS2,3*, and many more (Supplementary Table 3).

However, none of these genes or QTL were located at the position of the haploblock as identified in this study, thus, it could be a new locus. Another haplotype on chr1A was in the proximity of QTL *Qyr.nwafu-1AL* at ~587 Mb, which was previously reported using 90K and 660K markers. However, no major gene has been

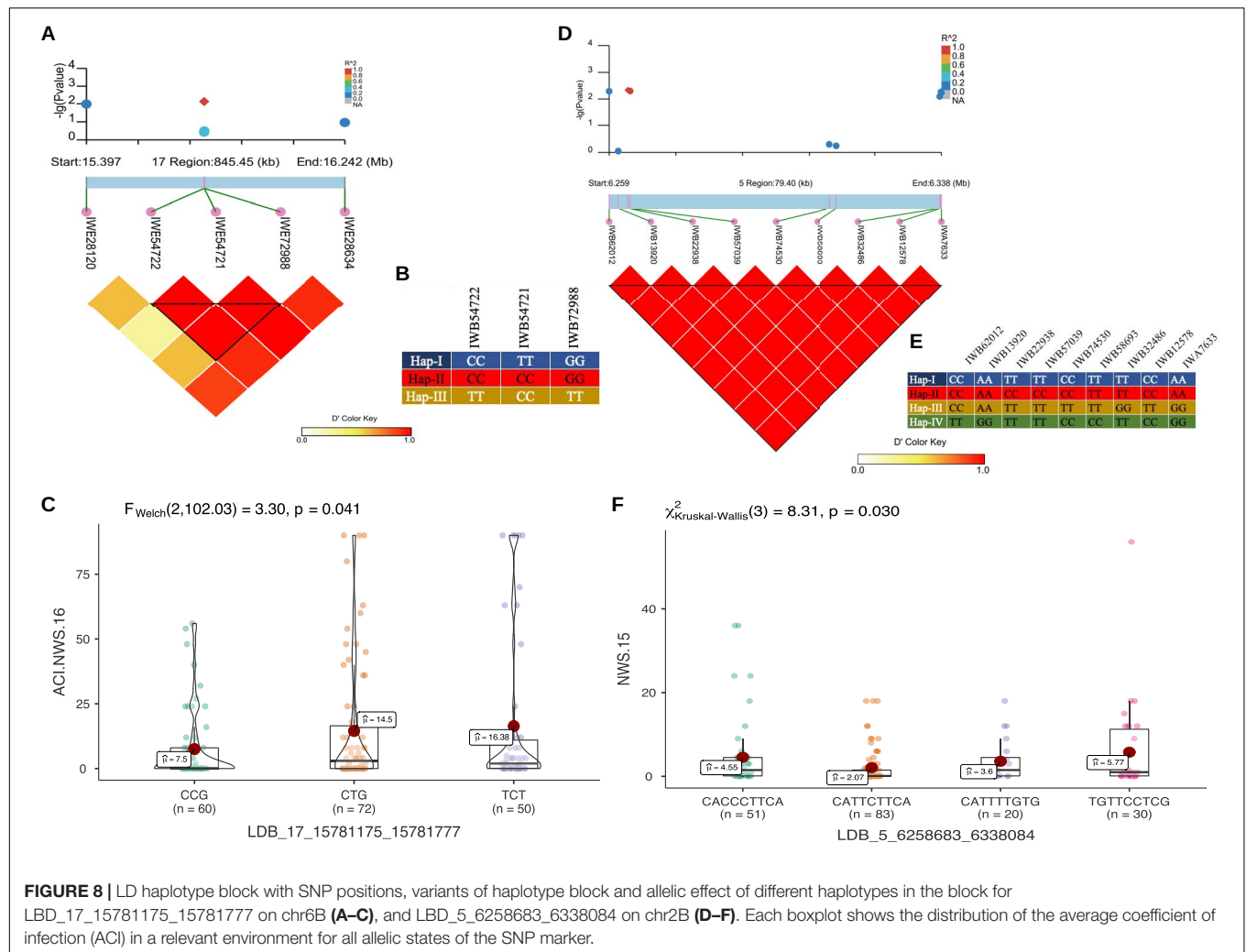


FIGURE 8 | LD haplotype block with SNP positions, variants of haplotype block and allelic effect of different haplotypes in the block for LBD_17_15781175_15781777 on chr6B (A–C), and LBD_5_6258683_6338084 on chr2B (D–F). Each boxplot shows the distribution of the average coefficient of infection (ACI) in a relevant environment for all allelic states of the SNP marker.

identified in this region. In conclusion, haplotype-GWAS proved to be a useful approach in combination with SNP-GWAS to improve the discovery of resistance loci.

Genomic Prediction for Stripe Rust Resistance

The transition from phenotypic selection to marker-assisted selection, and now genome-wide selection, will allow breeders to improve the selection decision during the early filial generations. However, the success of genomic selection depends on several factors such as the heritability of the trait, phenotypic variation explained by markers, and appropriate genomic prediction models (Ali et al., 2020). Genomic prediction resulted in an accuracy of up to 85% for APR at ISB.16, although predictions at other locations were less accurate. Prediction accuracies were low to moderate for the three other environments using the 90K SNP array but higher for GBS markers. The reason for low prediction accuracies can be attributed to smaller population sizes and unrelated genotypes. Recently, the prediction accuracy for stripe rust resistance in wheat landraces from Afghanistan was observed to be between 0.33

to 0.38 (Tehseen et al., 2021). Among the prediction models used, GBLUP and BayesB were the most effective, while EN was the least. The results in this study supported previous genomic prediction studies, where GBLUP and similar models predicted the disease resistance more accurately than other models (Avni et al., 2017; Juliana et al., 2017; Tehseen et al., 2021).

CONCLUSION

There is an ongoing need to identify new sources of resistance to *Pst*. The SYN-DERs provide valuable genetic resources for wheat improvement because they have high breeding value and are derived from primary synthetic hexaploidy wheats with D-genome contribution from *Ae. Tauschii*. Thus, SYN-DERs can be used to enhance the diversity of the D-genome in modern bread wheat but also the diversity of the A and B genomes because the synthetic wheats carry introgressions from durum wheat. More than 65 loci were identified in this study, which represent potentially important genes for race-specific and broad-spectrum

TABLE 7 | Genomic prediction accuracy using 11 different models for stripe rust resistance at four locations, and against five isolates at seedling stage using 90K SNP array and genotyping-by-sequencing (GBS) platform.

Markers		NWS.15	ISB.15	NWS.16	ISB.16	Pst.571242	Pst.571262	Pst.140202	Pst.571243
90K	BayesA	0.506 (0.052)	0.405 (0.053)	0.482 (0.07)	0.506 (0.038)	0.279 (0.113)	0.219 (0.074)	0.11 (0.05)	0.426 (0.059)
	BayesB	0.489 (0.055)	0.407 (0.05)	0.486 (0.07)	0.51 (0.044)	0.286 (0.111)	0.228 (0.079)	0.101 (0.051)	0.414 (0.062)
	BayesC	0.491 (0.061)	0.413 (0.052)	0.477 (0.077)	0.5 (0.044)	0.29 (0.104)	0.236 (0.076)	0.107 (0.055)	0.434 (0.061)
	BRR	0.488 (0.055)	0.395 (0.056)	0.502 (0.061)	0.498 (0.044)	0.264 (0.102)	0.236 (0.079)	0.139 (0.041)	0.413 (0.06)
	BL	0.511 (0.053)	0.392 (0.047)	0.47 (0.078)	0.493 (0.043)	0.275 (0.101)	0.24 (0.073)	0.102 (0.05)	0.42 (0.059)
	GBLUP	0.468 (0.058)	0.393 (0.048)	0.472 (0.069)	0.494 (0.044)	0.259 (0.103)	0.241 (0.077)	0.106 (0.038)	0.396 (0.059)
	RKHS	0.354 (0.054)	0.235 (0.054)	0.39 (0.081)	0.386 (0.062)	0.205 (0.094)	0.221 (0.064)	0.061 (0.049)	0.226 (0.084)
	EN	0.48 (0.062)	0.413 (0.06)	0.482 (0.069)	0.491 (0.043)	0.227 (0.104)	0.214 (0.085)	0.117 (0.033)	0.402 (0.056)
	RVM	0.505 (0.067)	0.396 (0.072)	0.486 (0.069)	0.466 (0.049)	0.245 (0.108)	0.212 (0.06)	0.131 (0.043)	0.34 (0.054)
	GP	0.476 (0.058)	0.392 (0.06)	0.5 (0.058)	0.498 (0.048)	0.258 (0.107)	0.25 (0.076)	0.088 (0.038)	0.408 (0.057)
GBS	RRBLUP	0.481 (0.057)	0.406 (0.052)	0.486 (0.069)	0.501 (0.042)	0.228 (0.099)	0.222 (0.078)	0.139 (0.035)	0.405 (0.058)
	BayesA	0.449 (0.108)	0.442 (0.077)	0.391 (0.061)	0.399 (0.089)	0.168 (0.047)	0.102 (0.07)	0.079 (0.053)	0.23 (0.051)
	BayesB	0.421 (0.118)	0.432 (0.083)	0.386 (0.06)	0.405 (0.086)	0.146 (0.046)	0.117 (0.067)	0.067 (0.056)	0.227 (0.048)
	BayesC	0.421 (0.114)	0.412 (0.084)	0.395 (0.061)	0.396 (0.084)	0.157 (0.053)	0.122 (0.07)	0.097 (0.053)	0.236 (0.044)
	BRR	0.407 (0.114)	0.426 (0.079)	0.398 (0.061)	0.428 (0.084)	0.146 (0.05)	0.106 (0.067)	0.062 (0.059)	0.229 (0.048)
	BL	0.428 (0.111)	0.399 (0.087)	0.38 (0.062)	0.391 (0.093)	0.15 (0.05)	0.109 (0.071)	0.054 (0.056)	0.241 (0.048)
	GBLUP	0.417 (0.111)	0.435 (0.081)	0.366 (0.069)	0.388 (0.086)	0.151 (0.051)	0.117 (0.069)	0.013 (0.054)	0.225 (0.05)
	RKHS	0.403 (0.115)	0.424 (0.082)	0.36 (0.064)	0.407 (0.09)	0.104 (0.043)	0.107 (0.07)	0.039 (0.052)	0.228 (0.052)
	EN	0.26 (0.086)	0.33 (0.109)	0.17 (0.104)	0.379 (0.066)	0.256 (0.049)	0.002 (0.087)	-0.01 (0.07)	0.084 (0.065)
	RVM	0.486 (0.106)	0.381 (0.08)	0.346 (0.072)	0.439 (0.084)	0.06 (0.04)	0.038 (0.071)	0.171 (0.06)	0.224 (0.083)
	GP	0.413 (0.114)	0.427 (0.087)	0.372 (0.069)	0.466 (0.084)	0.139 (0.049)	0.133 (0.071)	0.048 (0.066)	0.246 (0.056)
	RRBLUP	0.398 (0.109)	0.421 (0.082)	0.388 (0.061)	0.385 (0.087)	0.079 (0.056)	0.104 (0.07)	0.001 (0.053)	0.213 (0.055)

Genomic prediction models: BayesA, BayesB, and BayesC. BRR, Bayesian ridge regression; BL, Bayesian least absolute shrinkage and selector operator; GBLUP, genomic best linear unbiased prediction; RKHS, reproducing kernel Hilbert spaces regression; EN, elastic net; RVM, relevance vector machine; GP, Gaussian processor; rrBLUP, ridge regression best linear unbiased prediction. The values in the parentheses are SDs of the prediction accuracies.

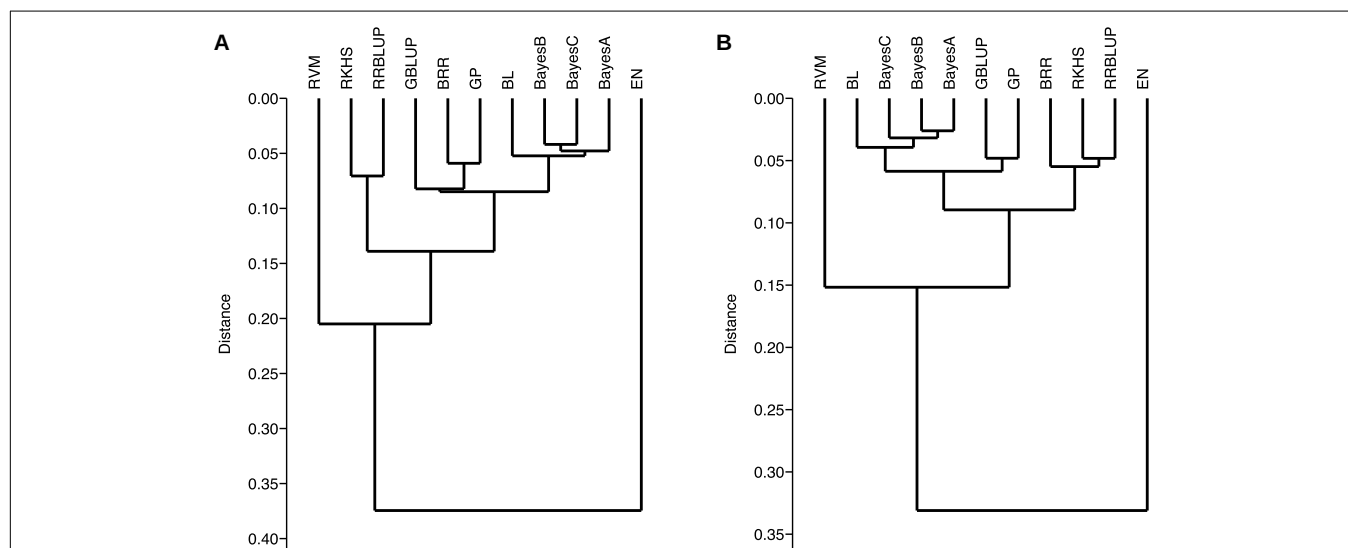


FIGURE 9 | Ward's hierarchical clustering on the prediction genomic values derived from the stripe rust infection types using 90K (A) and GBS (B) marker platforms. Genomic prediction models: BayesA, BayesB, BayesC, Bayesian ridge regression (BRR), Bayesian least absolute shrinkage and selector operator (BL), genomic best linear unbiased prediction (GBLUP), reproducing kernel Hilbert spaces regression (RKHS), elastic net (EN), relevance vector machine (RVM), Gaussian processor (GP), and ridge regression best linear unbiased prediction (rrBLUP).

resistance to stripe rust. Haplotype-GWAS should be a routine GWAS analytical approach to extend the discovery of genetic loci associated with phenotypes. The novel loci for resistance to stripe

rust identified by SNP, and haplotype GWAS provide an arsenal of new alleles for resistance breeding. The SNP markers with large phenotypic effects for both all-stage resistance and APR can be

converted to KASP or STARP markers for use in marker-assisted pre-breeding and breeding programs.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repository and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

ZM and AA performed field experiments. JM and MF performed the seedling experiment. AR, ZH, FO, and LH designed the experiment. RT conducted the 90K SNP genotyping. JP performed the GBS experiment. MA, AR, and KM analyzed

the data. ZM, KM, and AR wrote the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We acknowledge Pakistan Science Foundation (PSF) for financial support under PSF-NSFC III/Agr/C-QAU (16). We also acknowledge the National Natural Science Foundation of China (NSFC) for financial support under “Research Fund for International Young Scientist” (31950410563).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.788593/full#supplementary-material>

REFERENCES

- Afzal, F., Li, H., Gul, A., Subhani, A., Ali, A., Mujeeb-Kazi, A., et al. (2019). Genome-Wide Analyses Reveal Footprints of Divergent Selection and Drought Adaptive Traits in Synthetic-Derived Wheats. *G3* 9:1957. doi: 10.1534/g3.119.400010
- Ali, M., Zhang, Y., Rasheed, A., Wang, J., and Zhang, L. (2020). Genomic Prediction for Grain Yield and Yield-Related Traits in Chinese Winter Wheat. *Internat. J. Mole. Sci.* 21:1342. doi: 10.3390/ijms21041342
- Aoun, M., Rouse, M. N., Kolmer, J. A., Kumar, A., and Elias, E. M. (2021). Genome-Wide Association Studies Reveal All-Stage Rust Resistance Loci in Elite Durum Wheat Genotypes. *Front. Plant Sci.* 12:594. doi: 10.3389/fpls.2021.640739
- Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S. O., Gundlach, H., et al. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* 357, 93–97. doi: 10.1126/science.aan0032
- Azizinia, S., Bariana, H., Kolmer, J., Pasam, R., Bhavani, S., Chhetri, M., et al. (2020). Genomic prediction of rust resistance in tetraploid wheat under field and controlled environment conditions. *Agronomy* 10:1843.
- Babiker, E. M., Gordon, T. C., Chao, S., Newcomb, M., Rouse, M. N., Jin, Y., et al. (2015). Mapping resistance to the Ug99 race group of the stem rust pathogen in a spring wheat landrace. *Theoret. Appl. Genet.* 128, 605–612. doi: 10.1007/s00122-015-2456-6
- Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., and Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Sci.* 242, 23–36. doi: 10.1016/j.plantsci.2015.08.021
- Bhatta, M., Morgounov, A., Belamkar, V., Wegulo, N. S., Dababat, A. A., Erginbas-Orakci, G., et al. (2019). Genome-Wide Association Study for Multiple Biotic Stress Resistance in Synthetic Hexaploid Wheat. *Internat. J. Mole. Sci.* 2019:20. doi: 10.3390/ijms20153667
- Bulli, P., Zhang, J., Chao, S., Chen, X., and Pumphrey, M. (2016). Genetic Architecture of Resistance to Stripe Rust in a Global Winter Wheat Germplasm Collection. *G3* 6:2237. doi: 10.1534/g3.116.028407
- Chen, X. M. (2007). Challenges and solutions for stripe rust control in the United States. *Austral. J. Agric. Res.* 58, 648–655.
- Cox, T. S., Sears, R. G., Bequette, R. K., and Martin, T. J. (1995). Germplasm Enhancement in Winter-Wheat X *Triticum-Tauschii* Backcross Populations. *Crop Sci.* 35, 913–919.
- Crossa, J., Jarquin, D., Franco, J., Pérez-Rodríguez, P., Burgueño, J., Saint-Pierre, C., et al. (2016). Genomic Prediction of Gene Bank Wheat Landraces. *G3* 6, 1819–1834. doi: 10.1534/g3.116.029637
- Daetwyler, H. D., Bansal, U. K., Bariana, H. S., Hayden, M. J., and Hayes, B. J. (2014). Genomic prediction for rust resistance in diverse wheat landraces. *Theoret. Appl. Genet.* 127, 1795–1803. doi: 10.1007/s00122-014-2341-8
- Daetwyler, H. D., Villanueva, B., and Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3:e3395. doi: 10.1371/journal.pone.0003395
- Dreisigacker, S., Tiwari, R., and Sheoran, S. (2013). *Laboratory manual: ICAR-CIMMYT molecular breeding course in wheat*. Haryana: Directorate of Wheat Research Karnal, India.
- Duveiller, E., Singh, R. P., and Nicol, J. M. (2007). The challenges of maintaining wheat productivity: pests, diseases, and potential epidemics. *Euphytica* 157, 417–430.
- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4, 250–255.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33:1.
- Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., et al. (2002). The Structure of Haplotype Blocks in the Human Genome. *Science* 296:2225. doi: 10.1126/science.1069424
- Godoy, J. G., Rynearson, S., Chen, X., and Pumphrey, M. (2017). Genome-Wide Association Mapping of Loci for Resistance to Stripe Rust in North American Elite Spring Wheat Germplasm. *Phytopathology* 108, 234–245. doi: 10.1094/PHYTO-06-17-0195-R
- Habier, D., Fernando, R. L., and Garrick, D. J. (2013). Genomic BLUP Decoded: A Look into the Black Box of Genomic Prediction. *Genetics* 194:597. doi: 10.1534/genetics.113.152207
- Halloran, G. M., Ogonnaya, F. C., and Lagudah, E. S. (2008). *Triticum* (*Aegilops*) *tauschii* in the natural and artificial synthesis of hexaploid wheat. *Austral. J. Agric. Res.* 59, 475–490. doi: 10.1111/j.1365-3113X.2011.04667.x
- He, J., Meng, S., Zhao, T., Xing, G., Yang, S., Li, Y., et al. (2017). An innovative procedure of genome-wide association analysis fits studies on germplasm population and plant breeding. *Theoret. Appl. Genet.* 130, 2327–2343. doi: 10.1007/s00122-017-2962-9
- Huang, X., and Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* 65, 531–551. doi: 10.1146/annurev-arplant-050213-035715
- Jighly, A., Oyiga, B. C., Makdis, F., Nazari, K., Youssef, O., Tadesse, W., et al. (2015). Genome-wide DArT and SNP scan for QTL associated with resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in elite ICARDA wheat (*Triticum aestivum* L.) germplasm. *Theoret. Appl. Genet.* 128, 1277–1295. doi: 10.1007/s00122-015-2504-2
- Jordan, K. W., Wang, S., Lun, Y., Gardiner, L.-J., MacLachlan, R., Hucl, P., et al. (2015). A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biol.* 16:48. doi: 10.1186/s13059-015-0606-4
- Juliana, P., Singh, R. P., Singh, P. K., Crossa, J., Huerta-Espino, J., Lan, C., et al. (2017). Genomic and pedigree-based prediction for leaf, stem, and stripe rust

- resistance in wheat. *Theoret. Appl. Genet.* 130, 1415–1430. doi: 10.1007/s00122-017-2897-1
- Kankwatsa, P., Singh, D., Thomson, P. C., Babiker, E. M., Bonman, J. M., Newcomb, M., et al. (2017). Characterization and genome-wide association mapping of resistance to leaf rust, stem rust and stripe rust in a geographically diverse collection of spring wheat landraces. *Mole. Breed.* 37:113.
- Karatzoglou, A., Smola, A., Hornik, K., Karatzoglou, M. A., SparseM, S., and Yes, L. (2007). *The kernlab package. Kernel-Based Machine Learning Lab. R package version 0.9-22*. Available online: <https://cran.r-project.org/web/packages/kernlab> (accessed on 4 November 2015).
- Ledesma-Ramírez, L., Solís-Moya, E., Iturriaga, G., Sehgal, D., Reyes-Valdes, M. H., Montero-Tavera, V., et al. (2019). GWAS to Identify Genetic Loci for Resistance to Yellow Rust in Wheat Pre-Breeding Lines Derived From Diverse Exotic Crosses. *Front. Plant Sci.* 2019:10. doi: 10.3389/fpls.2019.01390
- Line, R. F., and Qayoum, A. (1992). *Virulence, aggressiveness, evolution and distribution of races of Puccinia striiformis (the cause of stripe rust of wheat) in North America, 1968-87*. Hyderabad: Dept. of Agriculture, Agricultural Research Service.
- Liu, W., Maccaferri, M., Ryneerson, S., Letta, T., Zegeye, H., Tuberosa, R., et al. (2017c). Novel Sources of Stripe Rust Resistance Identified by Genome-Wide Association Mapping in Ethiopian Durum Wheat (*Triticum turgidum* ssp. durum). *Front. Plant Sci.* 2017:8. doi: 10.3389/fpls.2017.00774
- Liu, W., Maccaferri, M., Chen, X., Laghetti, G., Pignone, D., Pumphrey, M., et al. (2017b). Genome-wide association mapping reveals a rich genetic architecture of stripe rust resistance loci in emmer wheat (*Triticum turgidum* ssp. dicoccum). *Theoret. Appl. Genet.* 130, 2249–2270. doi: 10.1007/s00122-017-2957-6
- Liu, W., Maccaferri, M., Bulli, P., Ryneerson, S., Tuberosa, R., Chen, X., et al. (2017a). Genome-wide association mapping for seedling and field resistance to *Puccinia striiformis* f. sp. tritici in elite durum wheat. *Theoret. Appl. Genet.* 130, 649–667. doi: 10.1007/s00122-016-2841-9
- Liu, W., Naruoka, Y., Miller, K., Garland-Campbell, K. A., and Carter, A. H. (2018). Characterizing and Validating Stripe Rust Resistance Loci in US Pacific Northwest Winter Wheat Accessions (*Triticum aestivum* L.) by Genome-wide Association and Linkage Mapping. *Plant Genome* 2018:11. doi: 10.3835/plantgenome2017.10.0087
- Long, L., Yao, F., Yu, C., Ye, X., Cheng, Y., Wang, Y., et al. (2019). Genome-Wide Association Study for Adult-Plant Resistance to Stripe Rust in Chinese Wheat Landraces (*Triticum aestivum* L.) From the Yellow and Huai River Valleys. *Front. Plant Sci.* 2019:10. doi: 10.3389/fpls.2019.00596
- Maccaferri, M., Zhang, J., Bulli, P., Abate, Z., Chao, S., Cantu, D., et al. (2015). A Genome-Wide Association Study of Resistance to Stripe Rust (*Puccinia striiformis* f. sp. tritici) in a Worldwide Collection of Hexaploid Spring Wheat (*Triticum aestivum* L.). *G3* 5:449.
- Manickavelu, A., Joukhadar, R., Jighly, A., Lan, C., Huerta-Espino, J., Stanikzai, A. S., et al. (2016). Genome wide association mapping of stripe rust resistance in Afghan wheat landraces. *Plant Sci.* 252, 222–229. doi: 10.1016/j.plantsci.2016.07.018
- McIntosh, R., Dubcovsky, J., Rogers, W., Morris, C., Appels, R., and Xia, X. (2016). Catalogue of gene symbols for wheat: 2015–2016 supplement. *Komugi Wheat Genet. Resour. Database*.
- McIntosh, R., Mu, J., Han, D., and Kang, Z. (2018). Wheat stripe rust resistance gene Yr24/Yr26: A retrospective review. *Crop J.* 6, 321–329. doi: 10.1016/j.cj.2018.02.001
- McNeal, F. H., Konzak, C. F., Smith, E. P., Tate, W. S., and Russell, T. S. (1971). *A Uniform System for Recording and Processing Cereal Research Data. Report No. REP-10904*. El Batán: CIMMYT.
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819
- Milus, E. A., Seyran, E., and McNew, R. (2006). Aggressiveness of *Puccinia striiformis* f. sp. tritici Isolates in the South-Central United States. *Plant Dis.* 90, 847–852. doi: 10.1094/PD-90-0847
- Mujeeb-Kazi, A., Rosas, V., and Roldan, S. (1996). Conservation of the genetic variation of *Triticum tauschii* (Coss.) Schmalh. (*Aegilops squarrosa* auct. non L.) in synthetic hexaploid wheats (*T. turgidum* L. s.lat. x *T. tauschii*; 2n=6x=42, AABBDD) and its potential utilization for wheat improvement. *Genet. Resour. Crop Evol.* 43, 129–134. doi: 10.1007/bf00126756
- Muleta, K. T., Bulli, P., Ryneerson, S., Chen, X., and Pumphrey, M. (2017a). Loci associated with resistance to stripe rust (*Puccinia striiformis* f. sp. tritici) in a core collection of spring wheat (*Triticum aestivum*). *PLoS One* 12:e0179087. doi: 10.1371/journal.pone.0179087
- Muleta, K. T., Rouse, M. N., Ryneerson, S., Chen, X., Buta, B. G., and Pumphrey, M. O. (2017b). Characterization of molecular diversity and genome-wide mapping of loci associated with resistance to stripe rust and stem rust in Ethiopian bread wheat accessions. *BMC Plant Biol.* 17:134. doi: 10.1186/s12870-017-1082-7
- Naruoka, Y., Garland-Campbell, K. A., and Carter, A. H. (2015). Genome-wide association mapping for stripe rust (*Puccinia striiformis* F. sp. tritici) in US Pacific Northwest winter wheat (*Triticum aestivum* L.). *Theoret. Appl. Genet.* 128, 1083–1101. doi: 10.1007/s00122-015-2492-2
- Nordborg, M., and Weigel, D. (2008). Next-generation genetics in plants. *Nature* 456:720. doi: 10.1038/nature07629
- Ogbonnaya, F. C., Abdalla, O., Mujeeb-Kazi, A., Alvina, G. K., Xu, S. S., Gosman, N., et al. (2013). Synthetic hexaploids: harnessing species of the primary gene pool for wheat improvement. *Plant Breed. Rev.* 37, 35–122. doi: 10.1002/9781118497869.ch2
- Ornella, L., Singh, S., Perez, P., Burgueño, J., Singh, R., Tapia, E., et al. (2012). Genomic prediction of genetic values for resistance to wheat rusts. *Plant Genome* 5, 136–148. doi: 10.1007/s00122-017-2897-1
- Pasam, R. K., Bansal, U., Daetwyler, H. D., Forrest, K. L., Wong, D., Petkowski, J., et al. (2017). Detection and validation of genomic regions associated with resistance to rust diseases in a worldwide hexaploid wheat landrace collection using BayesR and mixed linear model approaches. *Theor. Appl. Genet.* 130, 777–793. doi: 10.1007/s00122-016-2851-7
- Pérez, P., and De, G. L. C. (2014). Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* 198, 483–495.
- Rasheed, A., Wen, W., Gao, F. M., Zhai, S., Jin, H., Liu, J. D., et al. (2016). Development and validation of KASP assays for functional genes underpinning key economic traits in wheat. *Theor. Appl. Genet.* 129, 1843–1860. doi: 10.1007/s00122-016-2743-x
- Riaz, A., Athiyannan, N., Periyannan, S., Afanasenko, O., Mitrofanova, O., Aitken, E. A. B., et al. (2016). Mining Vavilov's treasure chest of wheat diversity for adult plant resistance to *Puccinia triticina*. *Plant Dis.* 101, 317–323. doi: 10.1094/PDIS-05-16-0614-RE
- Rutkoski, J. E., Poland, J. A., Singh, R. P., Huerta-Espino, J., Bhavani, S., Barbier, H., et al. (2014). Genomic Selection for Quantitative Adult Plant Stem Rust Resistance in Wheat. *Plant Genome* 2014:7.
- Solh, M., Nazari, K., Tadesse, W. and Wellings, C. (2012). "The growing threat of stripe rust worldwide," In *Borlaug Global Rust Initiative (BGRI) conference* (Beijing), 1-4.
- Tamba, C. L., and Zhang, Y.-M. (2018). A fast mrMLM algorithm for multi-locus genome-wide association studies. *bioRxiv* 2018:341784.
- Tehseen, M. M., Kehel, Z., Sansaloni, C. P., Lopes, M. D., Amri, A., Kurtulus, E., et al. (2021). Comparison of Genomic Prediction Methods for Yellow, Stem, and Leaf Rust Resistance in Wheat Landraces from Afghanistan. *Plants* 10:558. doi: 10.3390/plants10030558
- Turner, S. D. (2014). qqman: an R package for visualizing GWAS results using QQ and Manhattan plots. *Biorxiv*, 005165
- Vanraden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Wang, S. C., Wong, D. B., Forrest, K., Allen, A., Chao, S. M., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796. doi: 10.1111/pbi.12183
- Wang, S.-B., Feng, J.-Y., Ren, W.-L., Huang, B., Zhou, L., Wen, Y.-J., et al. (2016). Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* 6:19444. doi: 10.1038/srep19444
- Wellings, C. R. (2011). Global status of stripe rust: a review of historical and current threats. *Euphytica* 179, 129–141. doi: 10.1007/s10681-011-0360-y
- Wu, J., Yu, R., Wang, H., Zhou, C. E., Huang, S., Jiao, H., et al. (2021). A large-scale genomic association analysis identifies the candidate causal genes conferring stripe rust resistance under multiple field environments. *Plant Biotechnol. J.* 19, 177–191. doi: 10.1111/pbi.13452
- Yao, F., Long, L., Wang, Y., Duan, L., Zhao, X., Jiang, Y., et al. (2020). Population structure and genetic basis of the stripe rust resistance of 140 Chinese wheat

- landraces revealed by a sgenome-wide association study. *Plant Sci.* 301:110688. doi: 10.1016/j.plantsci.2020.110688
- Zegeye, H., Rasheed, A., Makdis, F., Badebo, A., and Ogbonnaya, F. C. (2014). Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. *PLoS One* 9:e105593. doi: 10.1371/journal.pone.0105593
- Zeng, Q. D., Han, D. J., Wang, Q. L., Yuan, F. P., Wu, J. H., Zhang, L., et al. (2014). Stripe rust resistance and genes in Chinese wheat cultivars and breeding lines. *Euphytica* 196, 271–284. doi: 10.1007/s10681-013-1030-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Mahmood, Ali, Mirza, Fayyaz, Majeed, Naeem, Aziz, Trethowan, Ogbonnaya, Poland, Quraishi, Hickey, Rasheed and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: frontiersin.org/about/contact



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership